2008

Metabolic engineering of Escherichia coli for the efficient utilization of plant sugar mixture

Madhuresh Kumar Choudhary
Iowa State University

Metabolic engineering of *Escherichia coli*

for the efficient utilization of plant sugar mixture

by

Madhuresh Kumar Choudhary

A thesis submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Major: Chemical Engineering

Program of Study Committee:
Jacqueline V. Shanks (Major Professor)
Monica H. Lamm
Julie A. Dickerson

Iowa State University

Ames, Iowa

2008

Table of contents

List of Figures

List of Tables

Abbreviations

ATP	Adenose Triphosphate
AC	adenylate cyclase
AMM	Atom mapping matrix
CCR	carbon catabolite repression
CLE	carbon labeling experiment
C-MFA	conventional metabolic flux analysis
cAMP	cyclic adenosien mono phosphate
ED	Empendroff
GCMS	Gas Chromatography Mass spectrometry
HPLC	High Performance Liquid Chromatography
IC	Information Content
IMM	Isotopomer mapping matrix
LDH	Lactose dehydrogenase
Mlc	making large colony
MFA	metabolic flux analysis
MetaFOR	Metabolic flux ratio
MG	Methylglyoxal
MC	Monte Carlo
NMR	Nuclear Magnetic Resonance
OD	Optical Density
OAA	oxaloacetate
PP	Pentose Pathway
PTS	phophotransferase system
PYR	Pyrvuate
SD	Standard Deviation

Abstract

In recent years, metabolic flux analysis (MFA) has become an important tool in metabolic engineering. The MFA involves the quantification of intracellular metabolic fluxes in a microorganism. The result of MFA is a metabolic flux map that allows the systematic study of cellular responses to genetic and environmental perturbations. Carbon isotope labeling based MFA, with analysis by NMR or GC/MS, have gained wide use in the metabolic engineering community for estimating metabolic fluxes in central carbon metabolism.. We are using MFA as a tool to engineer *Escherichia coli* sugar utilization regulatory (SURS) for the efficient consumption of sugar mixture.

E. coli SURS controls the utilization of different sugars as well as many other cellular functions. The SURS consist of the phosphoenolpyruvate (PEP)-dependent carbohydrate phosphotransferase system (PTS) and by several global regulators including CRP (cAMP receptor protein), Mlc (controlling several glycolytic, gluconeogenic and glucose-related genes), and Cra (catabolite repressor/activator protein). The tight control of sugar utilization in *E. coli* by SURS results in the sequential consumption of other sugars in the presence of glucose, which results in low yields and productivities of the desired product.

We have constructed *E. coli* strain devoid of gene encoding ptsG. The ptsG is involved in the transport of glucose and plays an important role in carbon catabolite repression. .The ptsG- strain didn't show diauxic growth and consumed glucose and xylose simultaneously. It grew slower than wild type grown on glucose and had a long lag time. Our flux analysis revealed that slow growth of ptsG mutant is due less efficient

transport of xylose and glucose in the mutant which results in less available ATP for biomass synthesis.

1 Introduction

The economy of the US and many other countries is dependent on the efficient supply of petroleum as energy and a source of raw materials for numerous chemicals. However, like other fossil fuels, it is not a renewable source of energy. The inadequate supply of oil can have severe consequence on US economy. In order to meet the growing demand for energy, there is a need for an environmentally sustainable source of energy. Plant biomass in the form of crops or agricultural waste is a potentially cheap and sustainable source of energy and raw material for many chemicals. The hydrolysis of cellulose from the plant biomass results in a mixture of fermentable sugars which can be converted in ethanol or other chemical through microbial fermentation. However, the consumption of pentose sugar is repressed in the presence of hexose sugar like glucose. In this work, we have used metabolic engineering approach to develop a strain of *Escherichia coli* capable of simultaneous consumption of both sugars at high rate.

Cellulose and hemicelluloses from plant biomass can be separated from the lignin and depolymerized to its constituent sugars, mainly glucose, xylose and arabinose. These fermentable sugars can be subsequently converted to ethanol and other chemicals by microbial fermentation. *E. coli* are one of the most suitable organisms for this purpose. It consumes these sugars at high rates, and grows robustly under industrial conditions with minimal nutritional requirements. It also has an established history as industrial "workhorse" for the production of wide variety of chemicals and therapeutic protein products [2].

The sugar mixture from plant biomass consists mainly of pentose and hexose sugars. The ability to utilize all sugars is a prerequisite for the efficient production of

ethanol or chemicals from the raw material. *E. coli* and several bacteria naturally possess a broad substrate-utilizing range. Nevertheless glucose prevents the simultaneous consumption of all sugars by catabolite repression effect (CCR)[3]. The fermentation of xylose and arabinose is delayed in the presence of glucose and is often incomplete, resulting in lower productivities and yield. The residual sugars are also problematic for downstream processing of products. Therefore, obtaining a recombinant strains capable of efficiently fermenting sugar mixture is a critical step in economical production of ethanol and other chemicals from plant sugar.

The phosphotransferase system (PTS) is the main system for glucose transport in *E. coli*. It consists of two cytosolic kinases (Enzyme I and HPr), and a sugar specific membrane bound protein (Enzyme II) that acts as a transporter. Furthermore PTS plays an important role in sugar uptake regulatory system (SURS) and catabolite repression. Several global regulators like CRP-cAMP, Mlc, and Cra are also involved in sugar uptake regulation. Hence, perturbation in SURS component could result in pleiotropic effects. The characterization of the system wide effects is necessary for our understanding of control of sugar metabolism in *E. coli*. To capture the effect on system wide effects require system biology tools like MFA and DNA microarray. *Metabolic engineering approaches are used in this work to attain our goal to create mutant capable of consuming both sugar at the same time*.

1.1 Metabolic Engineering and Metabolic Fluxes

Although, the concept of metabolic pathway manipulation for improving microbial and cellular process is an old concept [4], most of the earlier approaches are based on an empirical design which involves the creation of mutant library through

chemical mutagenesis and subsequent selection of the superior strain. Despite with many reported successes and wide spread acceptance the approach remains to be a random process where success is dependent in the “art “of selection of the mutant.

The development of molecular biological technique allows researcher to introduce precise genetic modification and led to the formation of the fields of genetic engineering and metabolic engineering [5, 6]. The metabolic engineering differs from genetic approach because of the emphasis it gives to system level approach. Genetic engineering considers a specific reaction or a metabolic pathway in isolation. Metabolic engineering on the other hand deals with interaction among various pathways. The analysis of precise genetic modification can give into the insight into future modification. This iterative approach of ME distinguishes from genetic engineering [7]..

Fundamental studies using genetic engineering approach have been carried out to evaluate the role of some of the individual components of SURS in the consumption of individual sugars and sugar mixtures [8-11]. However, most of these approaches were based on relating the effect of single mutations in individual components of the SURS to a “pooled’’ response (like growth rate, and sugar uptake rate). *Since the gene products interact in a complex, nonlinear and often totally unpredictable manner so that the simple genetic transformations are accompanied by unpredictable and often inexplicable results.* This approach fails to provide a tool to guide experiments for obtaining SURS mutant capable for efficiently metabolizing sugar mixtures. It is therefore, important to understand such interactions. In collaboration with Ramon Gonzalez and Ka-Yiu San at Rice University, we are going to use metabolic engineering strategies to meet this need. Unlike previous studies, our approach will be based on comprehensive modification and

characterization of the SURS in *E. coli*. First, the Gonzalez group will introduce precise genetic modification in single or multiple components of *E. coli* SURS. Then, we will evaluate global cellular responses to modifications through the quantification of metabolic fluxes (the Shanks research group) and evaluating the gene expression of individual genes through microarray (Gonzalez research group). The San research group will integrate transcriptional and flux data using a novel mathematical modeling framework of genetic network driven metabolic flux analysis which will guide us to determine the next set of genetic modifications to be introduced.

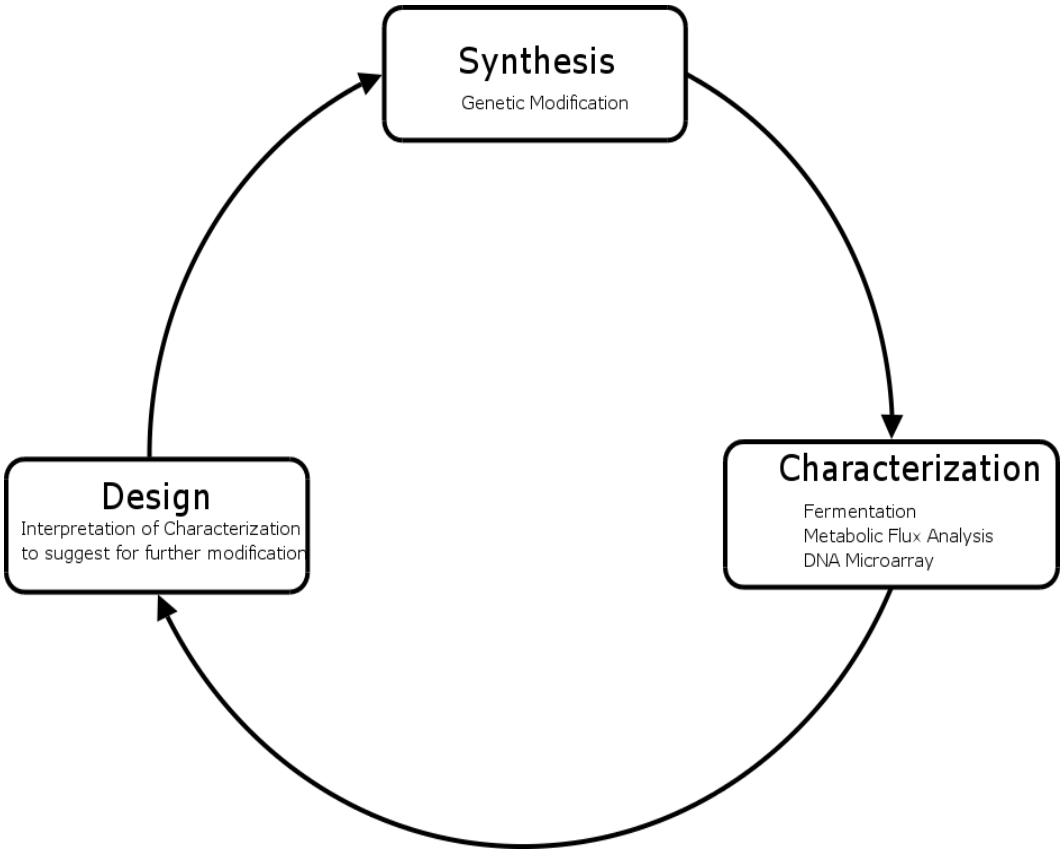

Figure 1.1: Metabolic engineering strategy for improving a microorganism. Genetic modification is followed by analysis. Metabolic flux analysis and transcript analysis via DNA microarrays are two modern tools to obtain insight for next round of modification.

1.2 Metabolic flux analysis

Metabolic flux analysis (MFA) has acquired a central place in metabolic engineering as a powerful diagnostic tool. MFA is the quantification of intracellular fluxes in a network. It provides information regarding the physiological state of the cell. Moreover, metabolic fluxes provide assessment of the impact of genetic modification and insight towards the selection of further metabolic engineering target.

Conventional flux analysis is the most basic form of metabolic flux analysis and is used as the first pass analysis tool. The key assumption is that the intracellular concentration of metabolite is constant at all times compared to the fluxes in and out of the metabolite, which is the pseudo-steady-state approximation. Conventional MFA is carried out by mass balance equations for intracellular metabolites and then solving the resulting set of linear equations. We have used conventional MFA to estimate metabolic fluxes for a simplified *E. coli* model.

As the metabolic network model becomes more complex, there are insufficient extracellular measurements to estimate metabolic fluxes. For example, conventional MFA can be used to estimate metabolic fluxes in case of a wild type *E. coli.* However, in case of the *ΔptsG*, one of *E. coli* mutants which will be used in this study, we hypothesize that the methylglyoxal pathway is active in contrast to the wild type strain. Hence, the metabolic model used in flux calculation for *ΔptsG* becomes more complex than that of wild type. Thus, extracellular measurements would be insufficient to estimate fluxes in *ΔptsG* strain.

Some of the limitations of the conventional flux analysis is overcome by supplementing it with ^{13}C labeling experiments. Each labeling measurement is reflective of intracellular fluxes and poses additional constraint on the metabolic network.^{13}C MFA

involves feeding a biological system with a combination of labeled and unlabeled substrates as carbon source. The biomass is then hydrolyzed to amino acids which are further analyzed by MS or NMR to determine their isotopomers distribution. The isotopomer distributions are in turn dependent on the intracellular fluxes. Intracellular fluxes are estimated from labeling measurement in iterative manner. Nevertheless, sometimes even external measurement and the NMR data are not sufficient to estimate fluxes through certain reactions, then the flux become 'unidentifiable'. This problem could be resolved either by increasing the number of measurements or with the proper choice of labeled substrate. While ^{13}C MFA of *E. coli* under aerobic conditions is well studied, anaerobic metabolism has been rarely studied, and ^{13}C MFA of microbes utilizing five and six carbon sugars has not been reported. We have used identifiability analysis for proper choice of labeled substrate and more of which will be discussed in Chapter 4.

The objectives of the overall collaborative project are: (i) to study the sugar regulatory systems (SURS) in *E. coli and* (ii) to engineer E. coli as biocatalyst that simultaneously consumes multiple sugars at high rate. The fundamental study will investigate the system wide effects of different mutations using MFA and DNA microarray, which hopefully will help us to elucidate catabolite repression phenomenon. Since, it is more economical to operate fermentor under anaerobic conditions and product yield is also higher under these condition, all studies will be performed under anaerobic conditions. The objective of this thesis is to (i) develop the flux analysis tool in *E. coli* for anaerobic metabolism and five and six carbon sugar consumption and (ii) to analyze flux distributions in wild-type E. coli and the *ΔptsG* strain.

1.3 Organization of the report

This chapter introduced the overall motivation for this work.

Chapter 2 gives the literature review of metabolic flux analysis.

Chapter 3 presents the literature review of *E. coli* sugar regulatory system and the need for the systematic approach.

Chapter 4 presents the results of anaerobic metabolic flux analysis. It also describes the design of future experiment by performing an identifiability analysis.

In Chapter 5, comparative study of wild type and *ΔptsG* mutant of *E. coli* by ^{13}C-MFA is presented.

Chapter 6 outlines the future experiment.

1.4 References

1. Campbell, C.J.L., J.H, *The end of cheap oil.* Scientific American, 1998. **3**: p. 78-83.
2. Gonzalez, R., et al., *Global Gene Expression Differences Associated with Changes in Glycolytic Flux and Growth Rate in Escherichia coli during the Fermentation of Glucose and Xylose.* Biotechnol. Prog., 2002. **18**(1): p. 6-20.
3. Milton H. Saier Jr., *Multiple mechanisms controlling carbon metabolism in bacteria.* Biotechnology and Bioengineering, 1998. **58**(2-3): p. 170-174.
4. Stephanopoulos, G., *Metabolic fluxes and metabolic engineering.* Metabolic Engineering, 1999. **1**(1): p. 1-11.
5. Bailey, J.E., *Toward a science of metabolic engineering.* Science, 1991. **252**(5013): p. 1668-1675.
6. Stephanopoulos, G., *Metabolic Engineering.* Current Opinion in Biotechnology, 1994. **5**: p. 169-200.
7. Stephanopoulos, G., J. Nielsen, and A.A. Aristidou, *Metabolic Engineering: Principles and Methodologies*. First ed. 1998: Elsevier Science & Technology Books.
8. Nichols, N.N., B.S. Dien, and R.J. Bothast, *Use of catabolite repression mutants for fermentation of sugar mixtures to ethanol.* Applied Microbiology and Biotechnology, 2001. **56**(1-2): p. 120-125.
9. Bothast, R.J., N.N. Nichols, and B.S. Dien, *Fermentations with New Recombinant Organisms.* Biotechnology Progress, 1999. **15**(5): p. 867-875.

10. Dien, B.S., N.N. Nichols, and R.J. Bothast, *Fermentation of sugar mixtures using Escherichia coli catabolite repression mutants engineered for production of L-lactic acid.* Journal of Industrial Microbiology & Biotechnology, 2002. **29**(5): p. 221-227.
11. Hernandez-Montalvo, V., et al., *Characterization of sugar mixtures utilization by an Escherichia coli mutant devoid of the phosphotransferase system.* Applied Microbiology and Biotechnology, 2001. **57**(1-2): p. 186-191.

2 Literature Review

2.5 Metabolic flux analysis

Cellular metabolism is driven by a large number of metabolic reactions. Central carbon metabolic pathways are involved not only in the conversion of the carbon source into building blocks needed for macromolecular biosynthesis, but also in the constant supply of Gibbs free energy via ATP and redox equivalents (NADPH/NADH) needed for the biosynthesis. As a result of evolution, the function of central carbon metabolism has been fine-tuned to exactly meet the needs for building blocks and Gibbs free energy in conjunction with cell growth rate. Therefore, metabolic fluxes i.e. rate of metabolic reaction, through the central carbon metabolism are tightly regulated[1].

The flux (V) through a given biochemical reaction can be specified as a function of two entities:

(1) the amount of enzyme(E) catalyzing the reaction;

and

(2) The concentration of the metabolites affecting enzyme activity, including the reactant and products of the enzyme reactions (Ci).

$$V = V(K's, E, Ci's) \tag{1}$$

$$\text{Specific Activity} = \frac{\text{Max Rate}}{\text{Amount of Enzyme}} \tag{2}$$

The flux through a reaction is a dependent on the metabolite level. Moreover, the concentration of the metabolites are themselves function of the metabolic fluxes and thus there is an important feedback regulation imposed on the system[1]. Additionally, the expression of enzyme is often regulated by the reactants and products of the enzymatic reaction. Thus, the metabolic flux is the final outcome of genetic, enzymatic and metabolic regulation and is a valuable representation of cell physiology. Flux measurements and comparisons of fluxes between different phenotypes can assist in the selection of appropriate metabolic engineering targets[2]. Hence, metabolic flux analysis (MFA) i.e., the quantification of intracellular metabolic fluxes in a pathway, is an important tool in metabolic engineering [3].

The rate of an "isolated" reaction can be estimated by measuring the rate of disappearance of product or rate of formation of reactant. However, the situation becomes complicated in case of network of reactions. The rate of change of reactant or product level is dependent on more than one reaction rate. Nevertheless, reaction rates can be estimated by measuring the level of each species involved in the reactions and using mass balance equations. In a cellular metabolism, various reactions of cellular metabolism are interconnected due to common metabolites, thus forming a large reaction network. Although, it is possible to estimate intracellular fluxes by measuring concentration of each metabolite, but it is problematic to measure intracellular concentration of each metabolite precisely. Therefore, intracellular metabolic fluxes are per se non-measurable quantities[2]. The one alternative to estimate metabolic fluxes is from the determination of *in vitro* enzyme kinetics. However, the results have to be treated with caution because assay conditions may not resemble the intracellular conditions. Another alternative is

based on the assumption of metabolic steady state or pseudo-steady-state. It is assumed that all fluxes coming into a given intracellular metabolite pool balance all fluxes going out of the pool. In pseudo-steady-state, the rate of change of intracellular metabolite concentration is small compared to fluxes in and out of the metabolite [1]. Basically, it implies that intracellular concentration of all metabolites can be assumed constant at all times. This is true for a cell growing at a physiological steady state.

Metabolic flux analysis can be performed using multiple approaches. In Figure 2.1, a systematic overview is given. The two most prominent of them are conventional flux analysis and ^{13}C based metabolic flux analysis.

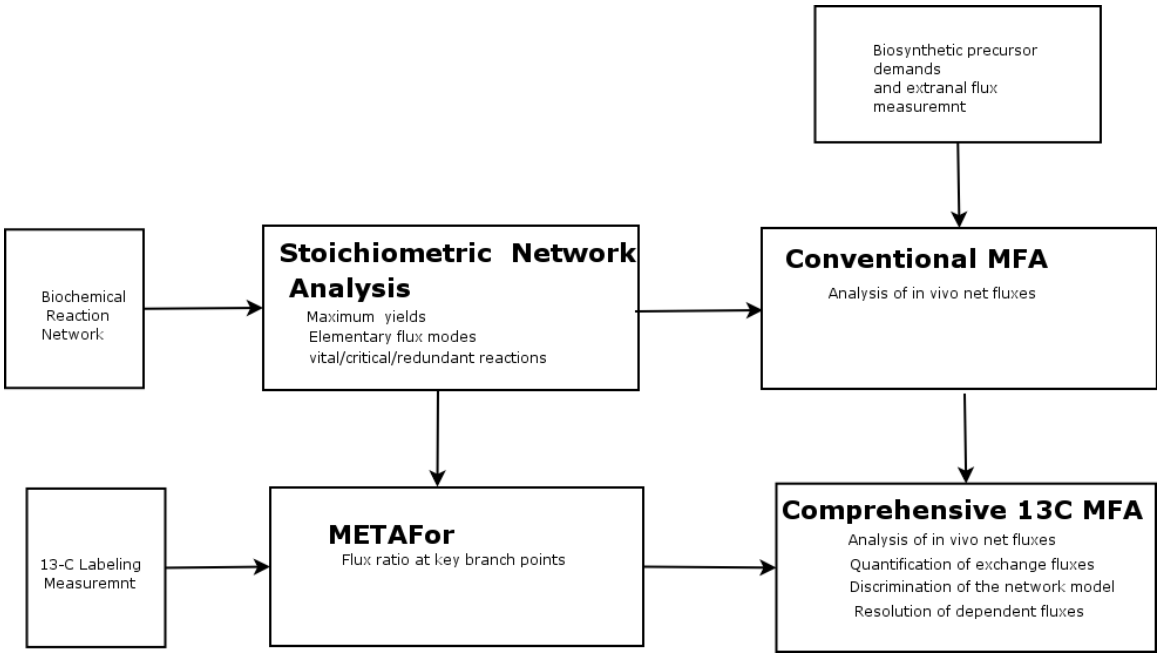

Figure 2.2: Different levels of metabolic flux analysis.

.

2.6 Conventional flux Analysis

The stoichiometric metabolic flux analysis (MFA) or conventional MFA is the most basic form of flux analysis and is based on metabolite balance around intracellular metabolites

with measurement of extracellular fluxes as constraints for flux calculation. Figure 3.2 illustrates the concept of metabolic flux for a very simple example.

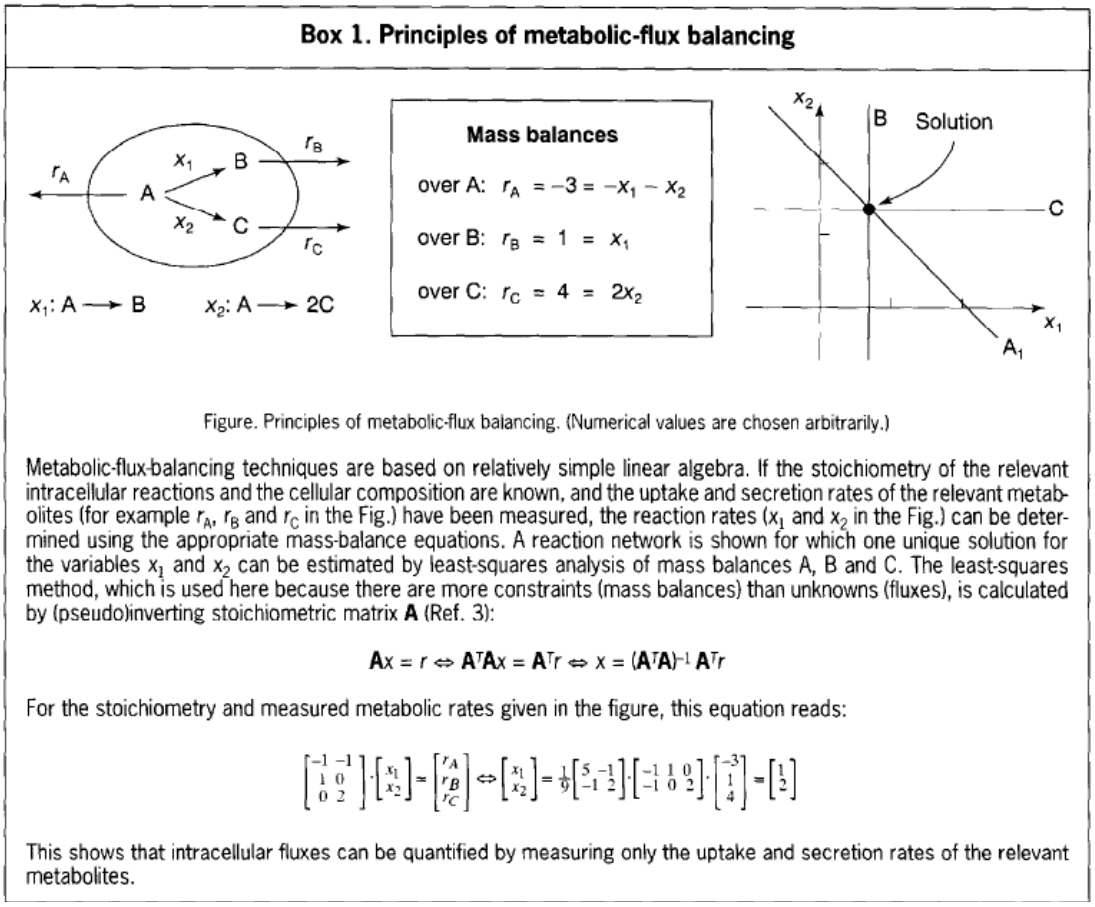

Figure 3.2: Principles of metabolic flux analysis. Figure is reproduced from [4]

In general, if there are J reactions and K internal metabolites, then the degree of freedom F=J-K. Through the measurement of F fluxes, the remaining fluxes can be calculated. If the number of supplied measurements is the same as F, it is a determined system; if greater than F, an over-determined system results and if less than F, it is an underdetermined system.

In case of an underdetermined system, additional constraints can be obtained by using NADH/NAPDH and ATP balances. But sometimes, incomplete pathway

knowledge can lead to erroneous flux estimation. Alternatively, optimization criteria like minimal ATP production or maximal growth rate are often used[2].

The stoichiometric flux analysis fails in case of parallel reactions and metabolic cycles[4]. The forward and back flux in reversible reactions cannot be resolved in stoichiometric MFA. Rarely in stoichiometric MFA, there are enough measurements to perform network validation.

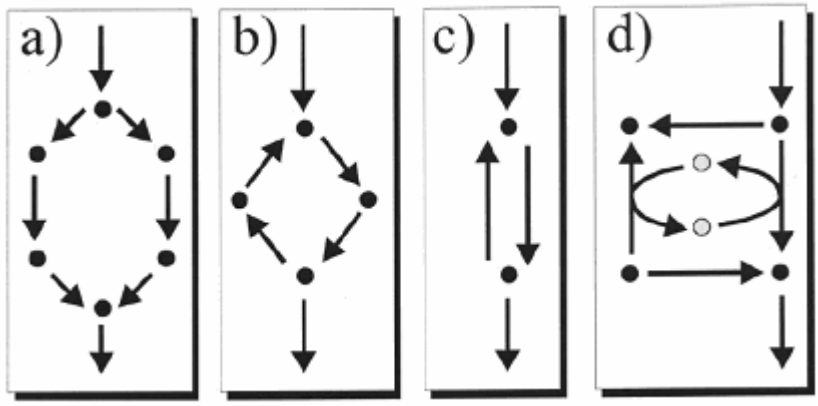

Figure 2.4: Typical situations in which stoichiometric MFA fails (a) parallel pathways (b) metabolic cycles (c) bidirectional reactions (d) split pathways when cofactors are not balanced. Figure reproduced from [[5]]

2.7 Carbon labeling experiment

The limitation of stoichiometric flux can be overcome by ^{13}C labeling experiments. Each labeling measurement that is dependent on the intracellular flux poses additional constraints on the set of intracellular fluxes[2]. In a stoichiometrically determined system, additional labeling measurements can be used to increase the statistical quality of the metabolic flux distributions as well as to validate network topology[2].

^{13}C MFA is based on a carbon-labeling experiment (CLE). In such an experiment, a ^{13}C -labeled substrate such as U-^{13}C glucose or 1-^{13}C glucose is fed to the biological system. The labeled carbon atoms are then distributed all over the metabolic network. The isotopomer distribution in the intracellular metabolite is dependent on the intracellular fluxes. The resulting data provides a large amount of additional information to quantify the intracellular fluxes[2]. Figure 2.5 summarizes the principle of ^{13}C MFA. From measured extracellular fluxes and measured labeling information the intracellular fluxes can be computed.

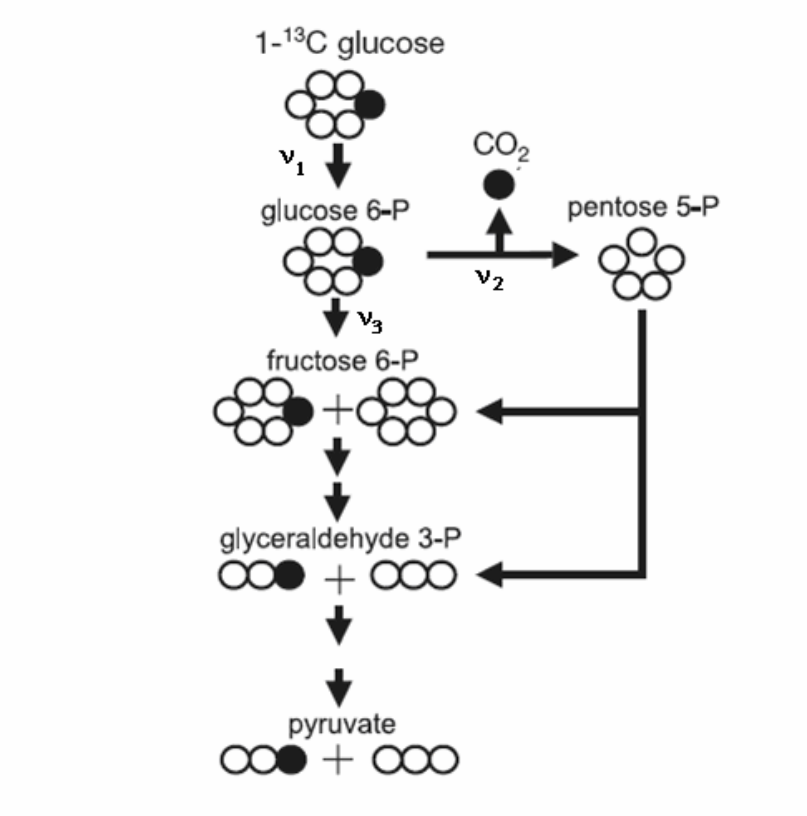

Figure 2.5: The ratio of 3-^{13}C Pyruvate and U-^{12}C Pyruvate is dependent on the flux ratio of pentose (v_2) and EMP (v_3) pathway**.** Figure adapted from [6].

A central concept of ^{13}C MFA is that of an isotopomer of a given metabolite[4]. The term isotopomer is a combination of the terms isotope and isomer and it means one of the possible different labeling states (Figure 2.6). Because a metabolite with 'n' carbon atoms can be labeled or unlabeled at each carbon atom position, there can be 2^n different labeling states of this molecule, which means that there are 2^n different isotopomers. Figure 2.6 shows the 2^3=8 different isotopomers of a metabolite with 3 carbon atoms.

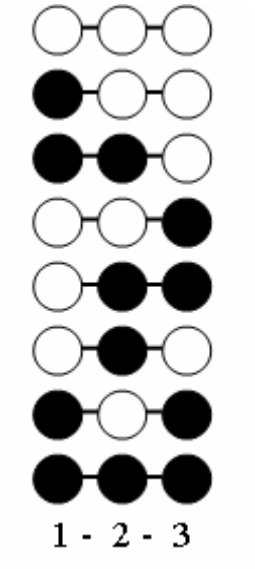

Figure 2.6: Isotopomer possibilities of three carbon metabolite. Filled and unfilled circles represent ^{13}C and ^{12}C atoms respectively.

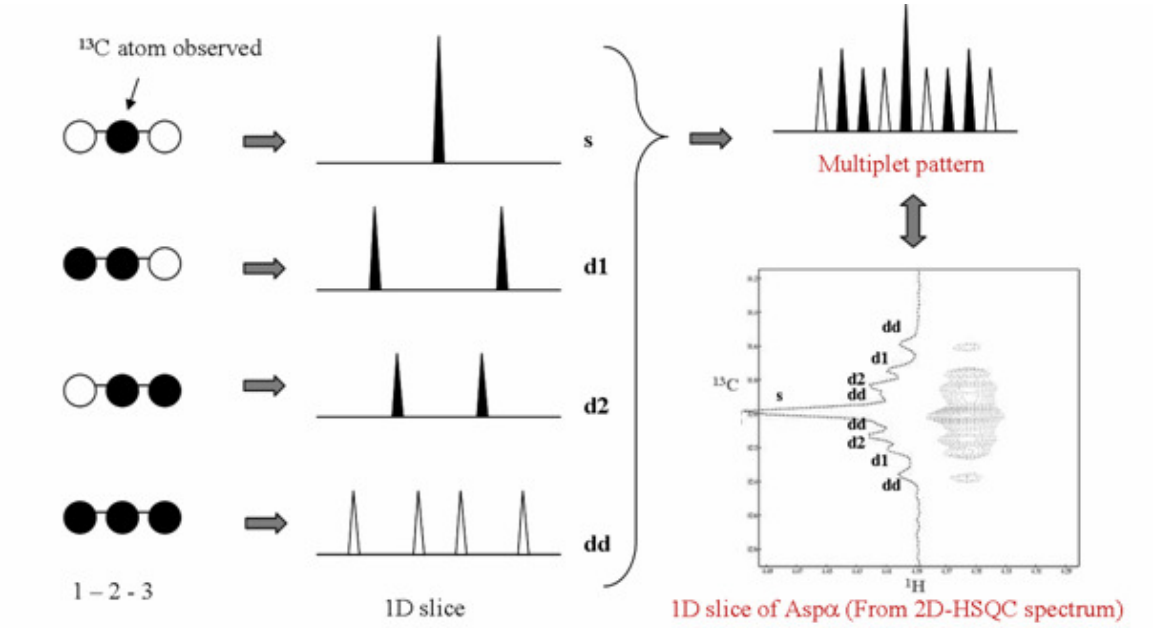

Figure 2.7: Relationship between isotopomer and multiplet pattern obtained from NMR measurement. Only 4 out of 8 isotopomers possible for three carbon metabolite are detected by NMR.

The isotopomers distribution of a metabolite with n carbon atoms is characterized by the percentage of each isotopomer within the metabolite pool, i.e., the isotopomer fractions. Clearly, the isotopomer fractions for each metabolite pool must sum up to one as illustrated in Figure 2.7. The isotopomer distributions of all metabolites are dependent on the metabolic fluxes through various pathways in the cell. Observing the intracellular labeling state of a microorganism in a CLE thus requires measured quantities that are related to the isotopomer distribution.

2.8 Measurement Procedures

Any measurement technique that can detect differences between isotopomers are suitable for gaining information about the intracellular labeling state of the system[6, 7].

Different types of NMR and MS techniques have been developed in order to obtain maximum information about the intracellular isotopomer distribution.

2.9 NMR techniques

NMR is a non-intrusive technique which can be used to detect labeling patterns of the metabolites. Proton NMR was the first method which was extensively applied to the ^{13}C-labeling experiment[8]. By this method each single protonated carbon atom position inside the particular metabolite pool can be observed separately from the other positions. The measured information is the positional enrichment of each carbon atom position.

In the ^{13}C NMR spectrum, the isotopomer distribution is resolved in more detail because a labeled carbon atom produces different hyperfine splitting signals depending on the labeling state of its direct neighbors in the molecule[9].

The information in a 1D spectrum is in the form of a number of peaks which may overlap and make the analysis difficult. So, the metabolites need to be isolated physically via separation technique for quantitative measurement. A 2D NMR correlation spectroscopy (COSY) is a combination of 1-H and 13-C NMR experiments and it detects the interaction of protons directly attached to carbon atom[10]. The advantage of this method as opposed to the formerly mentioned methods is that the different compounds do not have to be isolated from the hydrolyzates before actual measurement takes place[4]. In other words, the 2D NMR experiment performs the “separation”. A 2D HSQC (Heteronuclear Single Quantum Correlational spectroscopy) experiment can also be carried out to detect ^{13}C-^{13}C scalar couplings. In the HSQC experiment, the magnetization is transferred from the proton to the carbon and then back to proton. The indirect detection of carbon makes a HSQC experiment more sensitive than a COSY

experiment. In this study, 2D HSQC experiment has been used to detect the labeling pattern of carbon atoms in the amino acids from hydrolyzed biomass.

2.10 MS techniques review

NMR is a relatively insensitive technique and thus it require large amount of concentrated sample, typically between 10 - 40 mg of protein at 10% labeling, depending on the efficiency of label incorporation from the substrate. Moreover, NMR cannot detect all isotopomers (Figure 2.8). MS can used to get additional isotopomer measurements and it exhibits a much higher sensitivity than NMR[4]. In general, MS require less than $1/10^{th}$ of the sample required for NMR. There are multiple methodologies behind using MS. In GC-MS, the MS instrument is coupled to gas chromatogram to separate the compounds which are ionized, and fragmented and finally analyzed by MS[11]. Thus, not only mass isotopomer of the molecular ion are measured but also the mass isotopomer spectrum of several fragments. In LC-MS, the GC is replaced with a liquid chromatogram[4]. The main advantage of using a LC coupled with MS is that the chemical derivatization is not necessary as in the case of GC-MS. LC-MS can be time consuming and reproducible separation is not always possible. Recently, Pingitore *et al* demonstrated the use of Fourier transformation cyclotron resonance mass spectrometry (FT-ICR MS) to measure amino acid isotopomer distribution[12]. FTICR-MS is a type mass spectrometer for determining the mass-to-charge ratio (m/z) of ions based on the cyclotron frequency of the ions in a fixed magnetic field. The technique is fast and chemical derivatization is not required. However, structural isomers like leucine and isoleucine cannot be differentiated by this method [12].

For GC-MS measurement, the sample must be prepared by chemical derivatization. Additionally, to obtain ^{13}C labeling from GC-MS measurement, the data have to be corrected for natural labeling of analyte and the added derivatization residues [13]. Isotope effects have been also observed which means that the retention time of different labels depends on the isotopomer[4]. Similarly, overloading the MS detector must be prevented. MS like NMR cannot measure all isotopomers as shown in Figure 2.8

.

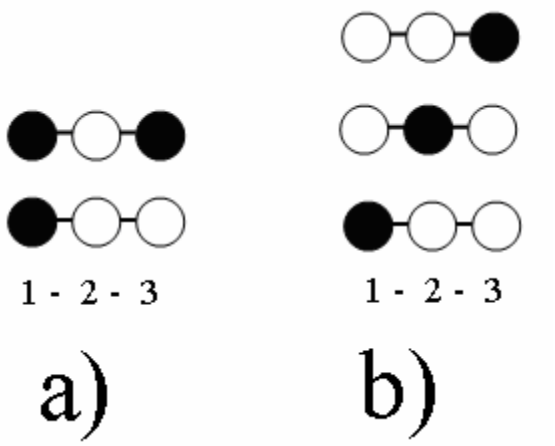

Figure 2.8: (a) NMR cannot distinguish between two isotopomers if only 1st carbon is observed **(b)** MS cannot distinguish above isotopomers

Recently, the numbers of reports of simultaneous use of GC-MS and NMR data along with biomass and extracellular measurement have been reported. The number of independent measurements increases from 47 and 57 using NMR spectroscopy and GC-MS, respectively, to 78 independent measurements using the combined approach of GC/MS and NMR[11].

2.11 Flux determination methodology

Since the relationship between the labeling state of biomass and intracellular flux distribution is nonlinear, it is impossible to find the analytical solutions for the intracellular fluxes[4]. However, labeling pattern of the biomass component can be calculated from the intracellular fluxes by simulating the carbon labeling experiment (CLE). Based on guessed flux values and the known input substrate composition, the steady state distribution of isotopomers over the network can be computed by simulating CLE. The set of fluxes that gives minimum deviation (χ^2) between experimental and simulated data is taken as an estimate of the real fluxes.

The mathematical procedure for simulating CLEs is computationally expensive. This comes from the fact that the mathematical model describing the dependency between the intracellular fluxes and the stationary isotopomer distribution contains one (often nonlinear) isotopomer balance equation for each isotopomer in the system. Katz and Wood [14] were the first who estimated the metabolic flux by solving the mass balance equation for positional enrichment of metabolite. However, this approach was only appropriate for small number of unidirectional fluxes. Zupke and Stephanopoulos[15] introduced the concept of Atom Mapping Matrices (AMMs) to numerically calculate the fractional enrichment of metabolites in a biochemical network [15]. However, the scheme cannot account for the labeling measurement provided by 2-D NMR. Schmidt [16] extended the AMM concept to isotopomer mapping matrices (IMMs). An iterative procedure was used to estimate isotopomers distribution at steady state. This made possible the simulation of all isotopomers in a metabolic network. However, the presence of large exchange fluxes causes severe instability of numerical solution. This restricted the application of this technique. Wiechert [17] found an elegant

way to overcome the problem of instability by reformulating isotope balance equations into cummomer balance equations. The term cummomer (cumulative isotopomer) designate the sum of isotopes which are labeled at a fixed position. Such reformulation made it possible to get analytical solutions of cummomer fractions by solving a cascade of linear equations starting with zero-order cummomer fractions. These cummomer fractions can be transformed to isotopomer fractions or used to calculate simulated NMR intensities.

To solve the optimization problem of minimizing the deviation between simulated and experimental data, generally a metaheuristics algorithm like Simulated Annealing (SA), Genetic Algorithm along with local optimization (Powell, simplex) are used[16]. To ensure that it has reached global optimum, simulations are done from different starting points to crosscheck if the program converges to the same optimum. Recently, Riascos reported the use of a deterministic algorithm(Branch and Bound) to estimate fluxes in *Saccharomyces cerevisiae* [18]. The advantage of a deterministic algorithm is that it ensures the global optimum but it is computationally expensive.

2.11.1 Statistical Analysis

Although, ^{13}C MFA have been successfully applied in various systems, the rigorous statistical analysis of estimated flux has received much less attention. It is often assumed that a large redundancy in the measurement set necessarily results in reliable estimates for fluxes. However, there have been reported instances when intracellular fluxes are associated with large standard deviations even in ^{13}CMFA. Since carbon labeling experiments are time consuming and expensive, therefore it is not practical to carry out the same experiment multiple times for the statistical analysis of the intracellular flux

distribution. Instead, computational methods are usually employed. Wiechert [19] reported that traditional linearized methods are not suitable for statistical analysis in the case of large exchange fluxes. The approximation quality of linearization is dependent on the curvature of linearized function. In case of high exchange fluxes, linear approximation is not able to follow the curvature of original function adequately. Hence, they proposed the use of compactification transformation by introducing the concept of reversibility for estimation of standard deviations (SD) by linearized methods. Since, linearized methods are based on a linear approximation of the nonlinear relation around the estimated flux, it may not give correct estimate of SD of fluxes. Additionally, it doesn't account for multiplicity of solution or non-ideal behavior (sub-optimal) of optimization techniques employed in flux estimation [2]. Hence, more rigorous analysis by Monte-Carlo (MC) Analysis has proposed by Schmidt et al[2]. In MC, a large number of synthetic experimental data is used as surrogates for multiple labeling experiments. These synthetic data are generated by adding random measurement errors to experimental data. The normal probability distribution of the NMR error is assumed for each data point. Monte Carlo simulation with the synthetic experimental data gives the probability distribution of the flux from which confidence intervals (CI) of the individual fluxes can be extracted. Although the MC approach is more accurate than the traditional linearized method, it is much more computationally expensive. Recently, Antoniewicz *et al* [20] compared the performance of method based on linearized statistics, Monte Carlo and grid search, which is even more rigorous than Monte Carlo. They found that CI estimated by the linearized method did not match with MC and grid search methods. On other hand, grid search and MC results agrees well with each other. Hence, MC gives correct

estimation of CI but it is computationally expensive. They reported the development of new method also based on linearized statistics but was shown to be accurate by comparing with MC and grid search.

Error associated with NMR or GC-MS is another area which has been ignored. Since NMR is not a sensitive technique, it is not possible to repeat the NMR experiment multiple times. Often errors of 0.01 is assumed or N/S ratio is assumed to be error associated with NMR intensities. However, this does not account for error due to overlapping of peaks and peak deconvolution and normalization of peaks. Van Winden[21] reported the development of a mathematical technique for estimation of error due to normalization of peaks. Later on[22], the method was extended to account for the error due to overlapping and deconvolution of peaks.

Compared to NMR, GC-MS is assumed to more accurate and can be repeated multiple times. From repeated runs in Christensen and Nielsen (1999), the measurement error of the MS instrument alone was deduced to be about 0.2% on the fractional enrichment scale. However, error due to overlapping of peak cannot be excluded by this technique. Often an error of 0.5% is assumed for MS measurement. Aljoscha *et al* reported the development of a tool which does error estimation based on noise associated with MS data in addition to natural isotope correction[23].

2.12 Flux Identifiability and the Design of optimal experiment

It is often not obvious a priori if the fluxes in a metabolic network can be estimated from the labeling measurement as labeling data are nonlinear functions of fluxes. Hence flux identifiability (whether flux can be estimated or not) is an important question in ^{13}C labeling based flux analysis. There are two aspects of flux identifiability:

1. Structural identifiability: The intracellular fluxes are called structurally identifiable if for any possible outcome of experiment and noise-free measurement there is a unique set of fluxes values producing the measured data.
2. Statistical identifiability: The intracellular fluxes are called statistical identifiable if fluxes can be estimated accurately from labeling measurement with noises.

Clearly, if the fluxes are not structurally identifiable, then they are also statistically unidentifiable. One example of structurally unidentifiable flux is the parallel reactions with same carbon rearrangement.

The flux identifiability is substantially altered by varying the labeling state of the substrate fed to the system, so there exists potential to design an 'optimal experiment' which best identifies the fluxes in a network. There are four choice of substrate:

(A) Unlabeled substrate;

(B) Specifically labeled substrate;

(C) Multiple specifically labeled substrate; and

(D) Uniformly labeled substrate.

A purely unlabeled as well as purely uniformly labeled substrate will yield no flux information at all as the labeling data obtained from such data are not dependent on fluxes but only on the type of labeling used. Thus, such substrates must always be mixed with another kind of substrate. Szyperski [10] and Schmidt *et al* [2] used different mixtures of uniformly labeled, unlabeled and 1-labeled glucose in *E. coli* and obtained good measurement of fluxes.

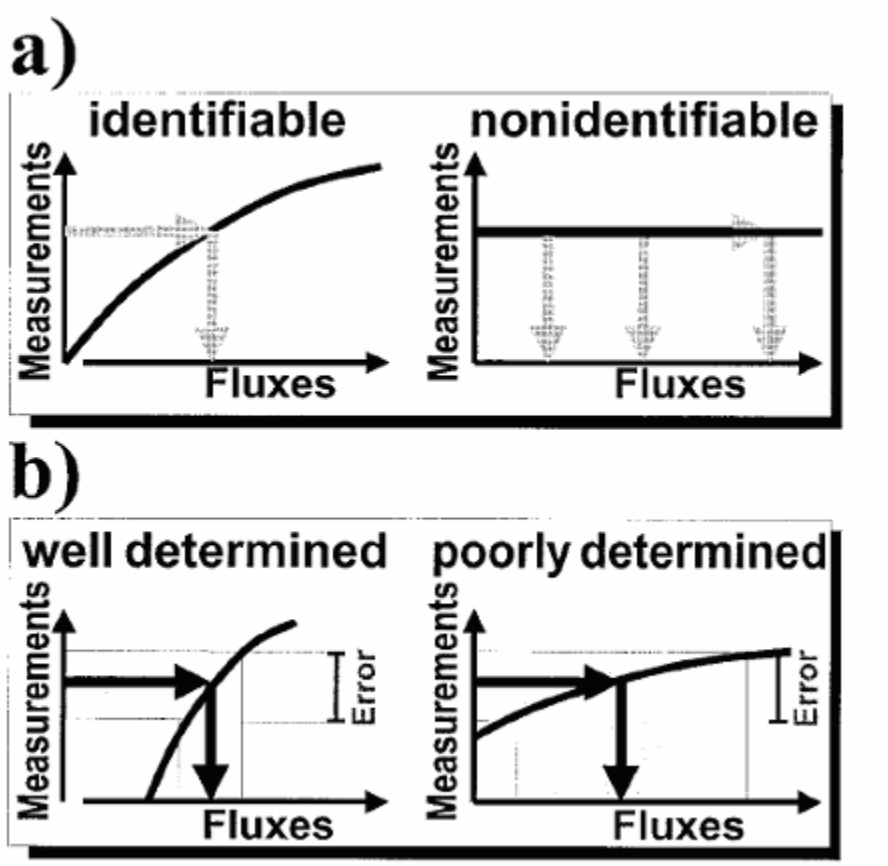

Figure2.9: One-Dimensional illustration of the flux identifiability problem. (a) Structural identifiability is given in the left figure but not in the right figure. (b) Even if fluxes are structurally identifiable they might not be identifiable in a statistical sense (left figure versus right figure). Figure reproduced from[5]

Szyperski [10] in his pioneering work, used 10% U-^{13}C glucose with 90% unlabeled glucose as the carbon source. Since labeling was used to calculate the bond integrity, they decided to use 10% U-^{13}C glucose by comparing the ratio of probability of two molecules being labeled in a fragmented molecule vs. unfragmented *molecule.* However, with development of MFA based isotope balancing, there was needed a new method for the optimal design of experiment. In *1999,* Wiechert developed a quantitative method to determine combination of substrate is best for a certain experiment [24]. The method uses linearized statistics method to calculated the SD of fluxes and geometrical

mean of the SDs is used as criteria as the performance of different experiments. They found the optimal substrate to study *E. coli* under aerobic conditions to consist of 1.00% unlabeled, 65.34% 1-^{13}C and 33.66% U-^{13}C glucose. In this work (Chapter 4), we examined the combinations of U-^{13}C glucose, 1-^{13}C glucose and unlabelled glucose by linearized method and followed by more rigorous analysis by Monte Carlo simulation.

However, another way to increase the accuracy of flux is by more accurate measurement of selected data. Stephanopoulos [20] developed a method to estimate the effect of change in a particular measurement on a flux estimate. The method can be used to identify the key measurement for a specific flux. Hence, suitable experiments can be carried out to measure them more precisely.

2.13 Labeling based metabolic flux analysis in Escherichia coli

Metabolic flux analysis has been used for over 20 years to explore cellular metabolism extensively for *E. coli*. Linear programming optimization with metabolite balances have been used to analyze metabolic networks[25]. Pramanik *et al* [26] created a comprehensive metabolic model for *E. coli* consisting of 300 reactions and 289 metabolites. The change in biomass composition with growth rate was also incorporated in the model. They concluded that the TCA cycle is not complete under anaerobic conditions and also that the Glyoxylate shunt pass is inactive under these conditions.

Another attempt to estimate metabolic flux in *E. coli* has been through conventional flux analysis [27]. The cells were grown on LB with glucose as carbon source and it was assumed carbon from glucose is not used for biomass synthesis but used only for the synthesis of fermentation products (Ethanol, Acetate, Lactate, and Succinate). Pentose pathway activity was neglected in their analysis as it was assumed

that pentose pathway role is limited to supply of NADPH and precursors for the biomass. With these simplifications their model was stoichiometrically determined.

Though ^{13}C based MFA has been extensively used for *E. coli* growing under aerobic conditions but its application on *E. coli* grown under anaerobic conditions is rather limited. Extracellular flux measurements as well as labeling measurements are essential to estimate intracellular fluxes. Since under anaerobic conditions, carbon flux to biomass is small compared to aerobic conditions and the most of carbon (85%) from glucose goes to fermentation products, extracellular measurement plays an important role in estimating intracellular fluxes. Moreover, the TCA cycle is not complete under these conditions which leads to less rearrangement. The rearrangement is important in determining metabolic fluxes. Hence, the proper choice of labeled substrate is important for fluxes under anaerobic conditions.

Szyperski pioneered the use of uniformly labeled substrate to estimate metabolic flux ratios at important metabolic branch points[10]. The labeling experiments were carried under three conditions (a) aerobic (b) micro-aerobic and (c) anaerobic conditions. The NMR data were used calculate the bond integrity of precursor molecules which were in turn used to estimate flux ratios. TCA cycle was concluded as not complete under anaerobic conditions and flux through the pentose pathway is small. Similar to the method used by Szyperski, Sauer carried out "METAFor" analysis with *E. coli* for different genetic and environmental conditions including anaerobic conditions [28] . The TCA cycle was determined to be branched under anaerobic conditions. The pentose pathway was concluded to be active under these conditions. The first rigorous ^{13}C-MFA by parameter fitting was carried for *E. coli* in Schmidt *et al*[16]. They applied MFA

based on previously reported NMR data [10]. However, biomass and extracellular measurement were obtained from literature [29]. Their metabolic model consisted of EMP and pentose pathway and TCA cycle and fermentative reactions. They found low flux (4.3) through TCA cycle under anaerobic conditions compared to aerobic conditions which was 45.0. They also found a very high flux through pentose pathway (77.7) which was even higher than under aerobic conditions. The inconsistency between their flux values and those estimated by Szyperski[10] was attributed to the effect of reversibilities of the EMP and PP reactions which were ignored in the analysis by Szyperski.

^{13}C flux analysis is the most reliable method available to estimate intracellular fluxes. But, there is no reliable estimate of metabolic fluxes in the literature for *E. coli* grown under anaerobic conditions. In this project, we have designed the labeling experiment under anaerobic conditions and ^{13}C labeling MFA under anaerobic conditions was performed. The result of comparative study of the wild type and its *ptsG* mutant grown on sugar mixture by ^{13}C MFA is presented in chapter 5.

2.14 References:

1. Nielsen, J., *It Is All about Metabolic Fluxes.* J. Bacteriol., 2003. **185**(24): p. 7031-7035.
2. Schmidt, K., et al., *Quantification of intracellular metabolic fluxes from fractional enrichment ^{13}C-^{13}C coupling constraints on the isotopomer distribution in labeled biomass components.* Metabolic Engineering, 1999. **1**: p. 166-179.
3. Stephanopoulos, G., *Metabolic fluxes and metabolic engineering.* Metabolic Engineering, 1999. **1**: p. 1-11.
4. Wiechert, W., *Minireview: ^{13}C Metabolic flux analysis.* Metabolic Engineering, 2001. **3**: p. 195-206.
5. Wiechert, W., et al., *A universal framework for ^{13}C metabolic flux analysis.* Metabolic Engineering, 2001. **3**: p. 265-283.
6. Wittmann, C., *Metabolic Flux Analysis Using Mass Spectrometry*. Advances in Biochemical Engineering/Biotechnology : Tools and Applications of Biochemical Engineering Science. 2002. 39-64.

7. Szyperski, T., *^{13}C-NMR, MS and metabolic flux balancing in biotechnology research*, in *Quarterly Reviews of Biophysics*. 1998. p. 41-106.
8. Marx, A., et al., *Determination of the fluxes in the central metabolism of Corynebacterium glutamicum by nuclear magnetic resonance spectroscopy combined with metabolite balancing.* Biotechnology and Bioengineering, 1996. **49**: p. 111-129.
9. Schmidt, K., et al., *^{13}C tracer experiments and metabolite balancing for metabolic flux analysis: Comparing two approaches.* Biotechnology and Bioengineering, 1998. **58**: p. 254-257.
10. Szyperski, T., *Biosynthetically directed fractional ^{13}C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism.* European Journal of Biochemistry, 1995. **232**: p. 433-448.
11. Christensen, B. and J. Nielsen, *Isotopomer analysis using GC-MS.* Metabolic Engineering, 1999. **1**: p. 282-290.
12. Pingitore, F., et al., *Analysis of Amino Acid Isotopomers Using FT-ICR MS.* Anal. Chem., 2007. **79**(6): p. 2483-2490.
13. Wittmann, C., *Fluxome analysis using GC-MS.* Microbial Cell Factories, 2007. **6**(1): p. 6.
14. Katz, J. and H.G. Wood, *Use of glucose-C14 for the evaluation of the pathways of glucose metabolism.* Journal of Biological Chemistry, 1960. **235**: p. 2165-77.
15. Zupke, C. and G. Stephanopoulos, *Modeling of isotope distributions and intracellular fluxes in metabolic networks using atom mapping matrices.* Biotechnology Progress, 1994. **10**: p. 489-498.
16. Schmidt, K., J. Nielsen, and J. Villadsen, *Quantitative analysis of metabolite fluxes in Escherichia coli, using two-dimensional NMR spectroscopy and complete isotopomer models.* Journal of Biotechnology, 1999. **71**: p. 175-190.
17. Wiechert, W., et al., *Bidirectional reaction steps in metabolic networks:III Explicit solution and analysis of isotopomer labeling systems.* Biotechnology and Bioengineering, 1999a. **66**: p. 69-85.
18. Riascos, C.A.M., A.K. Gombert, and J.M. Pinto, *A global optimization approach for metabolic flux analysis based on labeling balances.* Computers & Chemical Engineering, 2005. **29**(3): p. 447-458.
19. Wiechert, W., et al., *Bidirectional reaction steps in metabolic networks: II Flux estimation and statistical analysis.* Biotechnology and Bioengineering, 1997b. **55**: p. 118-135.
20. Antoniewicz, M.R., J.K. Kelleher, and G. Stephanopoulos, *Determination of confidence intervals of metabolic fluxes estimated from stable isotope measurements.* Metabolic Engineering, 2006. **8**(4): p. 324-337.
21. van Winden, W., P. Verheijen, and S. Heijnen, *Possible Pitfalls of Flux Calculations Based on 13C-Labeling.* Metabolic Engineering, 2001. **3**(2): p. 151-162.
22. van Winden, W., et al., *Innovations in Generation and Analysis of 2D [13C,1H] COSY NMR Spectra for Metabolic Flux Analysis Purposes.* Metabolic Engineering, 2001. **3**(4): p. 322-343.
23. S.Aljoscha Wahl, M.D.W.W., *New tools for mass isotopomer data evaluation in ^{13}Cflux analysis: Mass isotope correction, data consistency checking, and*

precursor relationships. Biotechnology and Bioengineering, 2004. **85**(3): p. 259-268.

24. Michael Möllney, W.W.D.K.A.A.d.G., *Bidirectional reaction steps in metabolic networks: IV. Optimal design of isotopomer labeling experiments.* Biotechnology and Bioengineering, 1999. **66**(2): p. 86-103.
25. Varma, A. and B.O. Palsson, *Stoichiometric flux balance models quantitatively predict growth and metabolic by-product secretion in wild-type Escherichia coli W3110.* Appl. Environ. Microbiol., 1994. **60**(10): p. 3724-3731.
26. J. Pramanik, J.D.K., *Effect of <I>Escherichia coli</I> biomass composition on central metabolic fluxes predicted by a stoichiometric model.* Biotechnology and Bioengineering, 1998. **60**(2): p. 230-238.
27. Berrios-Rivera, S.J., et al., *Effect of Glucose Analog Supplementation on Metabolic Flux Distribution in Anaerobic Chemostat Cultures of Escherichia coli.* Metabolic Engineering, 2000. **2**(2): p. 149-154.
28. Sauer, U., et al., *Metabolic Flux Ratio Analysis of Genetic and Environmental Modulations of Escherichia coli Central Carbon Metabolism*, in *The Journal of Bacteriology*

J. Bacteriol. 1999. p. 6679-6688.

29. Ingraham, J.L., Maaloe, O. and Neidhardt, F.C.,, *Growth of the Bacterial Cell*. 1983, Sunderland, MA.: Sinauer Associates, .

3 Sugar Utilization Regulatory System in *Escherichia coli*

3.1 Introduction

Lignocellulose from plant biomass is a potential raw material for fuels and chemicals. The hydrolysis of lignocellulose yields a mixture of sugars comprising of glucose, xylose and arabinose [1-3] *E. coli,* like most bacteria, is able to consume a variety of sugars. However, when exposed to a mixture of sugars, it selects the carbon source which supports a maximum growth rate. Therefore, they use sophisticated regulatory system that enables them to sense the nutrient situation and adjust their catabolic capacities by regulatory mechanisms which fall under the term carbon catabolism repression [4-6]. The carbon catabolite repression leads to the sequential and sometimes incomplete consumption of glucose, arabinose and xylose. Simultaneous consumption of sugars in a mixture would be advantageous in fermentative production process as it would eliminate growth in phases, therefore reducing operating time and increasing productivity[1]. The PTS system, which is the main transport system of glucose in *E. coli,* plays an important role in carbon catabolite repression (CCR)[7].

3.2 PTS system

The phosphotransferase system (PTS) is the major sugar transport system in many Gram-positive and Gram-negative bacterial species[7]. It is responsible for the entry of variety of carbohydrates in *E. coli* including glucose[5, 8, 9]. The glucose PTS catalyzes the following overall process:

PEP+ Glucose (out) → *Pyruvate +Glucose-6 Phosphate (in).*

The glucose phosphorylation is coupled to its translocation across the cytoplasmic membrane, the energy for the process is provided by the glycolytic intermediate phosphoenolpyruvate (PEP). Since glucose must be phosphorylated in order to enter the Embden-Meyerhof pathway (EMP) glycolytic pathway, it is the most efficient system as it consumes only one mole of phosphodiester bond for each mole of internalized phosphorylated glucose[5]. No unphosphorylated glucose can be detected inside the cells[7].

The glucose PTS comprised of two general proteins EI and HPr, which are common to all PTS, and two glucose specific proteins IIA^{glc} and $IIBC^{glc}$. These PTS proteins catalyze the following reactions[5, 9, 10]:

(1) PEP + enzyme I(EI) ←→ P-EI +Pyruvate

(2) P-EI+ HPr ←→P-HPr + EI

(3) P-HPr +$EIIA^{glc}$←→ P-$EIIA^{glc}$ + HPr

(4) P-$EIIBC^{glc}$ + Glucose (out) → $EIIBC^{glc}$+ Glucose-6-P (in)

The first step in the phosphotransferase reaction sequence catalyzed by the PTS is the phosphorylation of the general protein enzyme I (reaction 1). The phosphoryl group in phosphor-enzyme is linked to the N3 position in the imidazole ring of a histidine residue. The monomeric subunits of enzyme I are catalytically inactive. Dimers are formed in a monomer-dimer association reaction. Kinetic studies with the purified components indicate further that apparently only dimer can be phosphorylated.

Phosphoenzyme I and the soluble proteins of PTS have the highest phosphoryl transfer potential of all known naturally occurring phosphoryl compounds. Phospho-enzyme I can transfer its phosphoryl group back to Pyruvate to form PEP. The main function of EI

is to transfer its phosphoryl group to the second general protein of the PTS, the histidine containing phosphocarrier protein HPr to form phosphor-HP according to the reaction 2.

HPr, the second general PTS protein involved in the phosphoryl transfer reaction sequence is encoded by gene *crr*. It contains 84 amino acids with two histidine amino residues, at position 15 and 75. The phosphoryl group derived from phosphor-enzyme-I is attached solely to the N1-position of the imidazole ring of his-15. Thus, his-15 must be part of the active part of site. Interestingly, none of the enterobacterial HPr seems to contain tryptophan, tyrosine, or cysteine.

IIA^{glc} is glucose specific protein in *E. coli and it* acts as a phosphoryl carrier from phosphor-HPr to the membrane bound $IIBC^{Glc}$. The IIA^{glc} is also required for the uptake and phosphorylation of sucrose in *E. coli*. Similar to HPr, it lacks cysteine, tyrosine and tryptophan. Intact IIA^{glc} accepts one molecule of phosphate from PEP and is linked at the N3 position of imidazole ring of his-91, one of the two histidyl residue found in the molecule. Reaction 3 is completely reversible. The characteristics of IIA^{glc} are remarkable thermal stability and a strong tendency to form dimer, trimer or hexamers.

The $IIBC^{Glc}$ is a membrane bound protein, which forms a complex with IIA^{glc} and consist of two domains: EIIB, which is in contact with cytoplasm and EIIC, which is buried within the membrane. The phosphoryl group of EIIB is transferred to the glucose after translocation by IIC domain across the membrane and delivery at the face of the cytoplasmic membrane. As a general rule, enzymes IIBC have a broad specificity such that any substrate can be taken up by more than one enzyme II.

All the reactions up to and including EIIs are reversible. Only the final step (reaction 4), $P\sim EIIBC^{glc}$ + Glucose → Glucose-6-Phosphate + $EIIBC^{glc}$ is virtually

irreversible [11]. The reversibility of the phosphoryl group transfer in the PTS has regulatory consequences as it allows the metabolic network to the phosphorylation in a number of ways[10].

Tight pts mutations are unable to grow on pts carbohydrates if no other transport is available[12]. It has been suggested that $IIBC^{Glc}$ is closed pore, which is opened on phosphorylation and does not facilitate diffusion through its pore[13].

A mutant devoid of PTS cannot rapidly assimilate glucose and other PTS sugars; therefore, it displays an extremely low growth rate when glucose is used as carbon source[14]. This reduced but still existent, capacity to transport glucose depends on a low-capacity, high affinity system like that of LamB and Mgl [15]. Some degree of unspecificity of IIBC enzyme also allows transport of PTS sugars through different PTS. For example, Δ*ptsG* mutants (lacking $IIBC^{glc}$) can transport glucose through the II^{man} system, although with lower efficiency. Glucose can also be transported through the galactose proton symporter (GalP) and subsequent phosphorylation by glucokinase (Glk) constitute an alternative to PTS, as shown in Figure 3.1 [16].

3.3 Xylose and Arabinose utilization

Xylose is the most important five carbon sugar in the plant biomass and is the second most abundant sugar after glucose. It enters central carbon metabolism through the oxidative branch of the pentose phosphate pathway as xylose-5-phosphate[18]. The sugar is transported inside the cell through the high affinity system XylFGH, as shown in Figure 3., and a low affinity system XylE, a proton symporter. With sequencing, XylF was found as the xylose binding protein, xylG as ATP binding protein and XylH as a

membrane bound protein. Internalized xylose is isomerized into D-xylulose by xylose isomerase (XylA) and then phosphorylated by xylulokinase (XylB) to produce D-xylulose 5-phosphate [19]. The xylose genes are organized into two transcriptional units

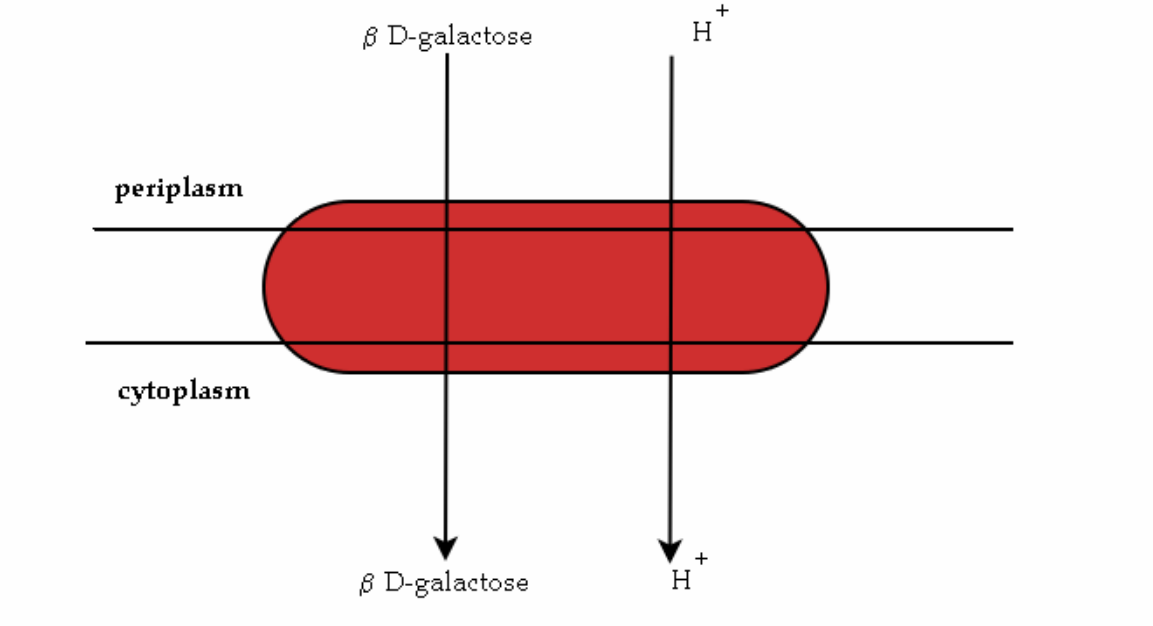

Figure 3.10: Galactose symporter GalP utilizes electrochemical gradients for galactose transport. It can also transport glucose. Figure adapted from Ecocyc[17]

which are linked but are oriented in opposite direction: (i) xylAB encoding xylose isomerase and xylulokinase under the P_A promoter, and (ii) xylFGHR encoding three subunits of the xylose ATP-binding cassette (ABC) transporter and transcriptional activator XylR[20]. XylR is a transcriptional activator for both promoters P_A and P_F, and is itself under a weak internal promoter, P_R. The transcription from P_A and P_F are induced by xylose and repressed by glucose. P_R is not regulated by these sugars and is under negative regulation by ArcA. The expression from promoter P_F is also under negative control of Fis and RpoS[21].

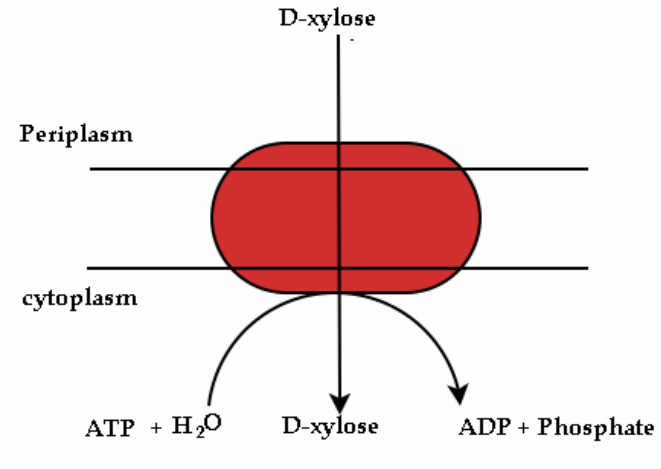

Figure 3.2: XylFGH is an ATP binding cassette (ABC) transporter which transports xylose with chemical potential of ATP.

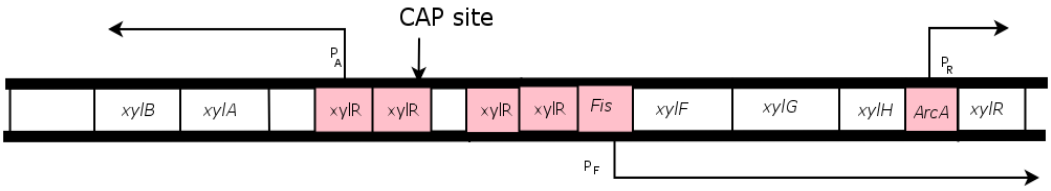

Figure 3.3: Organization of the xylose operon. The genes xylAB encodes for metabolic enzymes while xylFGH gene encodes for transport components. The *xylR* gene is involved in the transcriptional regulation of the *xyl* genes.

After xylose, arabinose is the next major five carbon sugar component of plant biomass accounting for 2 to 5% of total sugar present. The utilization of arabinose is similar to that of xylose [22]. Arabinose's dissimilation pathway starts with conversion to D-xylulose-5-phosphate, involving L-arabinose isomerase, L-ribulokinase, and L-ribulose-5-phosphate-4-epimerase. These enzymes are encoded by araBAD, which is analogous to xylAB. The high affinity transport system is encoded by araFGH, similar to xylFGH and low affinity symporters AraE and AraJ are analogous to XylE. The whole regulon (Figure 3.11) consists of five transcriptional units araBAD, araC, araFGH, araE

and araJ. The arabinose operon which is structurally similar to xyl operon is both negatively and positively regulated by AraC. In the absence of arabinose, araC dimers interact with the $araO_2$ site and $araI_1$ just upstream of P_{BAD} promoter, thus facilitating the formation of DNA loop and suppressing the transcription of araBAD [23-25]. When arabinose is present, the binding of monomeric araC shifts to $araO_1$, thereby activating the transcription of P_{BAD}. The CRP binds to a region upstream of araI, thereby opening the loop to initiate the transcription from P_{BAD} and P_C. Opening the loop may prevent the binding of AraC to $araI_1$ and $araO_2$. The close proximity of CRP-binding site to the AraC-binding site suggests interactions between AraC and CRP are involved in P_{BAD} activation.

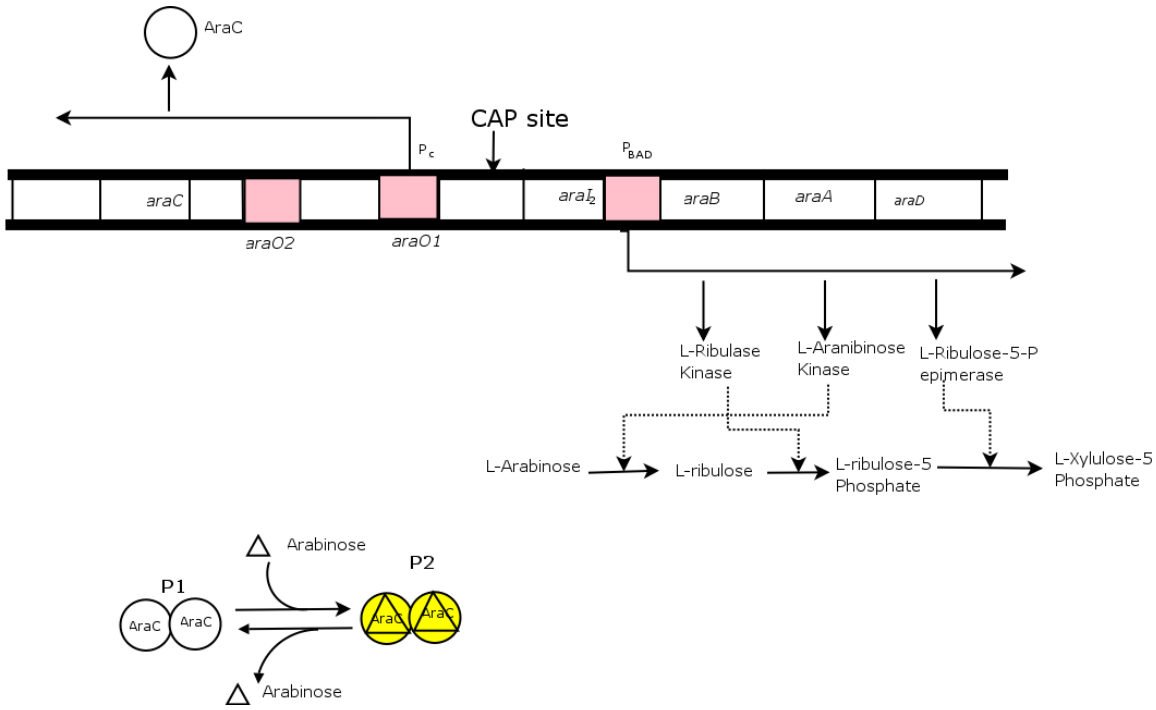

Figure 3.11: Organization of arabinose operon. The genes araBAD encoded for enzymes involved in arabinose degradation to xylulose-5-phosphate. The araC gene is involved in transcriptional regulation. AraC dimer binds to araO2 and araI site and thereby facilitating the formation of DNA loop. In presence of presence of arabinose, araC shifts to $araO_1$ from $araO_2$.

3.4 PTS as signal transduction mechanism

When bacteria are exposed to two different carbon sources, it used preferentially uses one and thus growth occurs in two phases[26]. As suggested originally by Monod, less preferred carbon sources are excluded from cells during the first phase by class A carbon source[7, 27]. For enteric bacteria, glucose is the best known substrate; hence the term "glucose effect" is used to describe the exclusion process.

A close connection between glucose uptake and the glucose effect had been always been noted. Most the mutants resistant against the glucose effect were found to be defective in glucose transport[7]. In *E. coli*, glucose is transported through PTS and its role in CCR became apparent when it was found that *ptsHI* mutants failed to grow not only on PTS substrates but also on non-PTS carbohydrates. Furthermore, the addition of cAMP in *ptsHI* mutant restores the growth on certain non-PTS compounds[8, 28]. The decisive clue to both to the central role of the PTS in the regulation of transport systems and catabolic pathways and to molecular mechanism involved was provided by a new type of suppressor mutant[13]. This suppressor mutation allowed the growth of pts mutants on non-pts sugars. The suppressor was called crr (catabolite repressor resistant). Later on, crr was shown to be structural gene for the IIA^{Glc}[29]. These results indicated that not only $EIIA^{glc}$ is involved in the glucose transport and phosphorylation but it also regulates the transport and metabolism of non-PTS sugars[3, 26, 30].

Based on these and other studies[31-33], a model (Figure 3.12) was proposed in which the protein IIA^{glc} in its phosphorylated form activates the enzyme adenylate cyclase. Adenylate cyclase synthesizes cAMP from ATP. The cAMP is the co-regulator of the cAMP-receptor protein CRP (also CAP), a global regulator for carbohydrate fluxes

in enteric bacteria. The cAMP-CRP complex interacts with the σ-subunit of the RNA-polymerase and acts as an activator for the transcription of nearly all genes responsible for the uptake and metabolism of carbohydrates, e.g. xylose or lactose. The amount of phosphorylated IIA^{glc}, is low when any PTS-carbohydrate like glucose or sucrose, is transported into the cell as Enzyme I-dependent transport is coupled necessarily to the phosphorylation of the transported substrates, and hence the dephosphorylation of the PTS-proteins including IIA^{glc} .Therefore, glucose reduces the cAMP level in the cell by decreasing the level of phosphorylated IIA^{glc} and thereby depriving the cell of CRP-cAMP level necessary for transcription of genes involved in the catabolism of other carbohydrates.

The cAMP mediated catabolite repression is not limited to glucose. The dephosphorylation of IIA^{glc} is also caused by transport of other PTS substrates, by competitive phosphoryl transfer from HPr-P. Non-PTS substrate (e.g. glucose-6-phosphate) induce catabolite repression by decreasing the intracellular PEP/Pyruvate ratio[34, 35]. CRP-cAMP also reduces the phosphorylation state of IIA^{glc}, the mechanism of which may be mediated by some other proteins that have not been identified as of yet. The possibility of E-I and HPr –mediated phosphorylation of IIA^{glc} by acetate kinase and ATP has been proposed [36].

The cAMP-mediated catabolite repression are related to the intracellular pool of cAMP (and thus activated CAP), any mechanism controlling this pool must be relevant to the understanding of this highly complex phenomenon. The internal pool of cAMP depends not only on its rate of synthesis but also on its rate of breakdown by phosphodiesterase and on efflux from the cells by an energy-dependent process that is

stimulated by metabolizable carbon sources. Exogenous glucose both inhibits the synthesis of cAMP and stimulates the efflux of cAMP from the cell cytoplasm.

Unphosphorylated enzyme IIA^{glc} can bind to a number of enzymes involved in the metabolism of non-PTS carbon sources and thereby inhibits the utilization of these carbon sources - a process called inducer exclusion. For example, IIA^{glc} binds to lactose transporter LacY, thereby preventing the formation allolacose that relieves the repression of the lac operon by binding to the repressor lacI. Only non-phosphorylated IIA^{glc} binds to the target protein; the phosphorylation prevents it's binding. Binding of IIA^{glc} occurs only when a substrate of target protein i.e. glycerol, glycerol or maltose is present. In this way, IIA^{glc} in the cell is not wasted on nonproductive binding which suggest that a conformational change in the target protein is required before the IIA^{Glc} can bind to it.

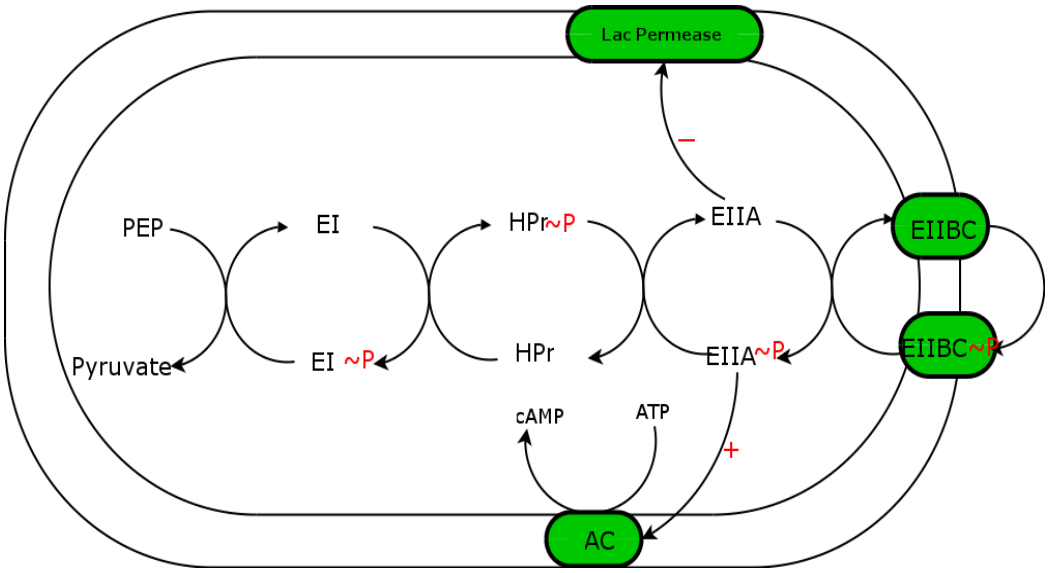

Figure 3.12: Model for the regulation by the PTS. In addition to phosphoryl transfer agent in glucose transport, unphosphorylated IIA^{glc} inhibits permease causing inducer exclusion while phosphorylated IIA^{glc} activates adenylate cyclase. Glucose transport through PTS leads to dephosphorylation of IIA^{glc} resulting in CCR either by inducer exclusion or by cAMP dependent catabolite repression.

The connection between glucose uptake rate, the phosphorylation state of $EIIA^{glc}$, adenylate cyclase activity and intracellular concentrations of PEP and pyruvate relate metabolic flux directly to the regulation of solute uptake and gene expression[10].

3.5 Regulation of PTS genes

The structural genes encoding for common PTS proteins ptsH, and ptsI are clustered along with the *crr* gene for glucose specific protein IIA^{glc} in a *pts* operon while glucose specific ptsG is expressed from distinct operon. Both *pts* and *ptsG* operon are induced during growth on glucose and was shown to require functional CRP-cAMP. Anaerobic conditions seem to favor high expression of pts proteins.

The transcription of pts genes is regulated from six promoters: p0a, p0b, p1a, p1b, px and p2[37]. The transcription of *ptsH* and *ptsI* is initiated mainly from promoter p0 and p1 while gene *crr* is expressed from promoter p2. The transcription of *crr* from p2 is not affected drastically either by CRP-cAMP or glucose mediated induction. However, the transcription from p1 and p2 is regulated in a complex way. The p0a and p1b promoter are activated by CRP-cAMP while p1a is repressed by CAP. In addition to regulation by CAP, p0a and p1b is repressed by Mlc and Cra respectively. Mlc binds to promoter region of P0 and preventing the binding of RNA polymerase. In the absence of glucose, the transcription is initiated from p1a and p0a while in the presence of glucose, repression by Mlc and Cra is relieved, it switches to p1b and p0a.

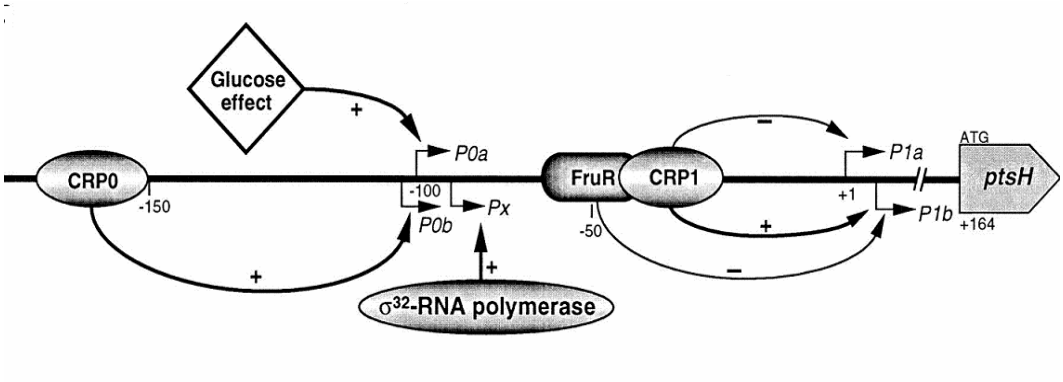

Figure 3.13: Schematic diagram showing the regulatory circuit of promoters of the *pts* operon by CRP-cAMP and FruR. Figure reproduced from [37].

ptsG transcription is initiated from two distinct promoters and both are regulated positively by CRP-cAMP, and negatively by global repressor Mlc. CRP-cAMP dependent expression from promoter p1 account for 90% of the total ptsG mRNA. In *cya* and *crp* mutants, the ptsG gene is not expressed.

$IIBC^{glc}$ gene expression is also affected by other factors. Anaerobic regulator (ArcA) binds to a region overlapping the binding site of P1 promoter, thereby negatively controlling *ptsG* transcription in response to the redox condition of the growth in *E. coli*..

Glucose induction of pts and ptsG operon is mediated by mlc. Unphosphorylated IIA^{glc} binds to Mlc and is sequestered away from its operator. Hence, glucose transport by PTS system leads to dephosphorylation of IIA^{glc} and thereby it leads to induction of mlc repressed genes. Since CRP-cAMP is required for the transcription of ptsG expression, Mlc repressed genes are not inducible in a CRP-cAMP mutant.

3.6 Post-transcriptional regulation of *ptsG*

From mutation studies, Morita *et al* [38] found that a mutation in either pgi or pfkA (encoding phosphoglucose isomerase or phosphofructokinase, enzymes catalyzing first two reactions of glycolysis) leads to rapid degradation of *ptsG* mRNA in a RNase E

dependent manner. Further studies revealed that RNase dependent destabilization occurs in response to accumulation of glucose-6-P or fructose-6-P[39]. Thus, *ptsG* expression is also regulated at the mRNA level in response to glycolytic flux. This would prevent unnecessary synthesis of ptsG and hence prevent the uptake of glucose when cell is no longer able to deal with it.

3.7 Adenylate Cyclase

Together with CRP, cAMP is involved in the regulation of a large number of catabolic genes. cAMP is synthesized from ATP by adenylate cyclase and the activity of adenylate cyclase is controlled by the PTS. Although there exist many *in vivo* results supporting the hypothesis that AC is activated by IIA^{glc}, it has never been demonstrated in vitro. Peterkofsky and coworkers[40] demonstrated that adenylate cyclase activity in toluenized *E. coli* cells is strongly inhibited by PTS carbohydrates provided the corresponding enzyme II is active. However, *in vitro* experiments with purified protein were not conclusive[40]. Most likely, other factors are also involved the activation of AC.

It has been observed repeatedly that *crp* mutants lacking CAP have increased levels of adenylate cyclase activity and secrete cAMP into the medium. Hence it is possible that CRP or another protein under negative control of CAP acts as repressor or inhibitor to the enzyme.

The *cya* gene encoding for adenylate cyclase is expressed from three promoters. CAP binds to the strongest promoter cyaP2 and represses the *cya* gene expression consistent with observed increased gene expression in *crp* mutant. Other two promoter may be responsive the binding of CAP. Thus, in cells growing under glucose-repressing conditions, cya gene expression is maximal. Increased *cya* gene expression in wild type

E. coli provides the bacteria the ability to produce large amounts of cAMP when the first carbon source is exhausted and thereby rapidly adapting to a new carbon source.

3.8 CRP

In *E. coli*, CRP is a global regulator involved in the transcription of more than 100 promoters. CRP needs the allosteric effector cAMP in order to bind to efficient to DNA. In fact, the expression of *crp* gene is itself under autoregulation by CRP-cAMP complex. Aiba *et al* [41] have shown that the cAMP-CRP complex both inhibits and activates the transcription of crp. The level of cAMP level required for activation is higher than those required for inhibition. Thus, in the presence of glucose (low cAMP), cellular CRP levels go down to enhance the intensity of repression and while in the absence it goes up to bring about the CRP mediated activation of catabolic operon.

3.9 CRA

The catabolite repressor/activator (Cra), initially characterized as fructose repressor (FruR), can bind upstream of certain promoters to activate their expression and can bind to sites overlapping other promoters to block transcription [42]. By using both DNA migration retardation assays and DNAse I footprint analyses, Cra (FruR) was found to bind to two operators within the regulatory region preceding the structural genes of fructose operon fruBKA[43]The *E .coli* FruR mutant exhibited pleiotropic phenotype, being unable to grow on lactate pyruvate and all Krebs cycle intermediates [43].

Since fructose phosphate interacts with FruR to prevent its binding to these sites, the presence of fructose or glucose presumably increases the intracellular concentration of the fructose phosphatase and consequently prevents the binding of Cra to its sites in

the promoter regions of target operons[42]. As a result glucose may partially block the initiation of transcription of promoters requiring activation by Cra, thus mediating catabolite repression and it promotes the initiation of transcription at promoters where Cra blocks the initiation of transcription. Because glycolytic enzymes and fermentative enzymes are repressed by Cra whereas gluconeogenic and oxidative enzymes are activated by Cra, glucose induces the former while repressing the latter[42].

3.10 Expression of Mlc

The mlc genes are weakly autoregulated; mlc is expressed from two promoters with an mlc site overlapping the downstream start site. A CAP site serves to regulate both promotes by repressing the upstream and activating the downstream site. The downstream promoter is recognized by RNA polymerase containing σ70 and the heat sigma factor σ32. Expression of mlc increases during heat shock but rapidly decays when glucose is present, implying mlc expression is under posttranscriptional control. In addition, *mlc* expression was shown to be enhanced under anaerobic conditions [44].

3.11 Autoregulation of Carbohydrate uptake

The sugar uptake regulatory system enables the bacteria to selectively utilize the carbon source which supports maximum growth rate and hence, it leads to sequential utilization of carbohydrates. Recently, Bruckner *et al* [45] proposed that it also helps bacteria to adjust sugar utilization to their metabolic capacities. In lactose grown cells, contrary to widespread belief, cAMP level is low, in fact, even lower than glucose grown cells. Additionally, inducer exclusion is still operational as half of $EIIA^{glc}$ exists in a non-phosphorylated form. A mutant with a lactose permease that is insensitive to inducer

exclusion exhibits a higher lactose uptake rate, but lower growth rate and *lac* expression than wild type[46]. Likewise, unregulated uptake of non-PTS sugar by an externally added phosphorylated sugar or cAMP was shown to cause cell death as a result of accumulation of methylglyoxal (MG), a toxic metabolite produced from dihydroxyacetone (DHAP) [47]. Hence, the global autoregulatory system is important for cells in adjusting the uptake rate to their metabolic capacity.

3.12 Engineering and utilization of PTS- strains

When *E. coli* is exposed to mixture of carbon sources, it will select the one that affords highest growth rate. As discussed previously, the transport of carbon source through the PTS leads to CCR by inducer exclusion and/or cAMP mediated repression. Different group have explored strategies to disrupt CCR by inactivating PTS components [2, 48, 49].

Fosfomycin, a analog of PEP, was used to select for the strain that was superior to strain KO11 (*E. coli* containing chromosomally integrated genes pdc_{Zm} and $adhB_{Zm}$, encoding the ethanol pathway from *Zymomonas mobilis*) for ethanol production from hexose and pentose sugars[49]. Fosfomycin resistant mutants do not use PTS to transport PTS sugars and usually they have a mutation in ptsI gene. These mutants displayed higher rates of sugar consumption when grown on a mixture containing 30g/l each of glucose, arabinose and xylose. Moreover, they produced 20% more ethanol than wild type with xylose as the sole carbon source.

Nichols *et al* [2] have studied the effect of *ptsG* inactivation on the pattern of sugar mixture utilization and its impact on ethanol production. The ptsG mutant and wild type were compared in cultures performed with minimal medium containing 2 g/l each of

either glucose and arabinose or glucose and xylose. The *ptsG* mutant consumed these sugars simultaneously whereas a wild type strain displayed sequential glucose-pentose consumption. In both conditions, the *ptsG* mutant consumed the total amount of sugars in about half the time of the wild type strain. A similar study was performed with a PTS^- Glc^+ strain obtained from a PTS^- Glc^- mutant by a continuous culture selection method [14]. When grown in a medium containing 1 g/l each of glucose, arabinose and xylose, the PTS^- Glc^+ strain consumed the total amount of sugars in the medium 16% faster than an isogenic PTS^+strain. Simultaneous consumption of glucose and arabinose was observed, but xylose consumption was delayed and but xylose consumption started when cells were still consuming glucose. Interestingly, the specific growth rate is reported to be $0.52hr^{-1}$ which is 21% higher than the wild type growing glucose or glucose/xylose mixture.

In another study, *ptsG* mutant was generated from the *E. coli* strain(pfl-,ldh-,ldhspt) engineered for L-lactic acids production[48]. In a culture performed with 50g/l of glucose and xylose, the ptsG mutant fermented three times more xylose than ptsG+ strain. Additionally, lactic yield was 1.5 times higher in ptsG- strain compared to ptsG+ strain.

Because of impaired transport system, PTS- mutants exhibit decreased sugar uptake rates and therefore slower growth. Snoep *et al* [50] showed the glucose utilization can be restored in PTS mutants through the expression of glucose facilitator protein Glf and Glk from *Zymomonas*. With the selection with PTS- strain in a continuous culture which has growth similar to wild type, it has been reported that galP gene in chromosome is necessary for rapid growth on glucose. Consequently, Hernandez *et al* [51] showed that

overexpression of alternative glucose transporters GalP and Glk restores growth rate and glycolytic flux to near wild type levels.

3.13 Proposed Research Approach

E. coli SURS is highly complex system. The global regulators of SURS like Cra, CAP, and Mlc modulate carbohydrates uptake and the direction of fluxes in central carbon metabolism by regulating the gene expression of various enzymes. Additionally, PEP/Pyr also link glycolysis with TCA/fermentative pathways. As mentioned previously, PEP/Pyr ratio enables bacteria regulate the carbohydrate uptake with their metabolic capacity. Hence, perturbation of PTS system will have impact on both glycolysis and TCA/fermentative pathways. One of the strategies to abolish CCR is the inactivation of PTS system. In the absence of PTS system, glucose is transported through less efficient transport system results in changes in intracellular level ATP, H^+ and NADH. Since ATP, NADH are involved in large number of reactions, effect of such modification will have profound effect on entire metabolic network. Therefore, tools that allow large-scale study of cellular network are needed. The collaborative efforts of Gonzalez, Shanks and San propose to use ^{13}C based MFA and genome wide transcriptomics analysis as tool to understand the system.

Metabolic engineering approach is used in this work which consists of three parts: synthesis, analysis and design. Synthesis part consists of genetic modification of SURS component *of E. coli*. Effect of genetic modification will be characterized by flux analysis (Dr. Shanks' group) and transcriptomics analysis (Dr. Gonzalez's group). Finally, the flux analysis and microarray will be

integrated using mathematic model which will give insight to design next of genetic modification. The comparative study of w3110 and *ptsG* mutant by ^{13}C MFA is presented in chapter 4. The DNA microarray analysis currently is being carried out in Dr. Gonzalez lab. The *ptsG* mutant consumed both glucose and xylose simultaneously however, at slower rate compared to wild type. Our flux analysis results indicate the sugar consumption could be improved with more efficient transport system for xylose and glucose in the *ptsG* mutant.

3.14 References

1. Bothast, R.J., N.N. Nichols, and B.S. Dien, *Fermentations with New Recombinant Organisms.* Biotechnol. Prog., 1999. **15**(5): p. 867-875.
2. Nichols, N.N., B.S. Dien, and R.J. Bothast, *Use of catabolite repression mutants for fermentation of sugar mixtures to ethanol.* Applied Microbiology and Biotechnology, 2001. **V56**(1): p. 120-125.
3. Tanaka, Y., et al., *Membrane localization itself but not binding to IICBGlc is directly responsible for the inactivation of the global repressor Mlc in Escherichia coli.* Molecular Microbiology, 2004. **53**(3): p. 941-951.
4. Titgemeyer, F. and W. Hillen, *Global control of sugar metabolism: a Gram-positive solution.* Antonie van Leeuwenhoek, 2002. **V82**(1): p. 59-71.
5. Postma, P.W.L., J.W.; Jacobson,G.R., *Phosphoenolpyruvate:carbohydrate phosphotransferase systems*, in *Escherichia coli and Salmonella:Cellular and Molecular Biology*, F.C. Niedhart, Schaechter,M.; Curtiss III,R; Lin,E.C.C; Low,K.B; Magasanik,B.; Reznikoff,W.S.;, Editor. 1996, American Society for Microbiology: Washington,D.C. p. 1149-1174.
6. Stulke, J. and W. Hillen, *Carbon catabolite repression in bacteria.* Current Opinion in Microbiology, 1999. **2**(2): p. 195-201.
7. Postma, P.W., J.W. Lengeler, and G.R. Jacobson, *Phosphoenolpyruvate:carbohydrate phosphotransferase systems of bacteria.*, in *Microbiol. Mol. Biol. Rev.* 1993. p. 543-594.
8. Saier, M.H., Jr ;Ramseir,T.M; Reizer,J., *Regulation of carbon utilization*, in *Escherichia coli and Salmonella:Cellular and Molecular Biology*, F.C. Niedhart, Schaechter,M.; Curtiss III,R; Lin,E.C.C; Low,K.B; Magasanik,B.; Reznikoff,W.S.;, Editor. 1996, American Society for Microbiology: Washington,D.C. p. 1325-1343.
9. Plumbridge, J., *Regulation of gene expression in the PTS in Escherichia coli: the role and interactions of Mlc.* Current Opinion in Microbiology, 2002. **5**(2): p. 187-193.

10. Deutscher, J., C. Francke, and P.W. Postma, *How Phosphotransferase System-Related Protein Phosphorylation Regulates Carbohydrate Metabolism in Bacteria.* Microbiol. Mol. Biol. Rev., 2006. **70**(4): p. 939-1031.
11. Meadow, N.D., et al., *Transient state kinetics of enzyme IICBGlc, a glucose transporter of the phosphoenolpyruvate phosphotransferase system of Escherichia coli: Equilibrium and second order rate constants for the glucose binding and phosphotransfer reactions.* Journal of Biological Chemistry, 2005. **280**(51): p. 41872-41880.
12. Postma, P.W., *Defective enzyme II-BGlc of the phosphoenolpyruvate: sugar phosphotransferase system leading to uncoupling of transport and phosphorylation in Salmonella typhimurium.* Journal of Bacteriology, 1981. **147**(2): p. 382-9.
13. Postma, P.W. and J.W. Lengeler, *Phosphoenolpyruvate:carbohydrate phosphotransferase system of bacteria.* Microbiol. Mol. Biol. Rev., 1985. **49**(3): p. 232-269.
14. Flores, N., et al., *Pathway engineering for the production of aromatic compounds in Escherichia coli.* Nature Biotechnology, 1996. **14**(5): p. 620-3.
15. Notley-McRobb, L. and T. Ferenci, *The generation of multiple co-existing mal-regulatory mutations through polygenic evolution in glucose-limited populations of Escherichia coli.* Environmental Microbiology, 1999. **1**(1): p. 45-52.
16. Flores, S., et al., *Analysis of carbon metabolism in Escherichia coli strains with an inactive phosphotransferase system by 13C labeling and NMR Spectroscopy.* Metabolic Engineering, 2002. **4**(2): p. 124-137.
17. Keseler, I.M., et al., *EcoCyc: a comprehensive database resource for Escherichia coli*
10.1093/nar/gki108. Nucl. Acids Res., 2005. **33**(suppl_1): p. D334-337.
18. Lin, E.C.C., *Dissimilatory pathways for sugars, polyols, and carboxylates.* Escherichia coli Salmonella typhimurium, 1987. **1**: p. 244-84.
19. David, J.D. and H. Wiesmeyer, *Control of xylose metabolism in Escherichia coli.* Biochimica et Biophysica Acta, General Subjects, 1970. **201**(3): p. 497-9.
20. Song, S. and C. Park, *Organization and regulation of the D-xylose operons in Escherichia coli K-12: XylR acts as a transcriptional activator.* Journal of Bacteriology, 1997. **179**(22): p. 7025-7032.
21. Xu, J. and R.C. Johnson, *Fis activates the RpoS-dependent stationary-phase expression of proP in Escherichia coli.* Journal of Bacteriology, 1995. **177**(18): p. 5222-31.
22. Ogden, S., et al., *The Escherichia coli L-arabinose operon: Binding sites of the regulatory proteins and a mechanism of positive and negative regulation.* Proceedings of the National Academy of Sciences of the United States of America, 1980. **77**(6): p. 3346-50.
23. Carra, J.H. and R.F. Schleif, *Variation of half-site organization and DNA looping by AraC protein.* EMBO Journal, 1993. **12**(1): p. 35-44.
24. Lobell, R.B. and R.F. Schleif, *AraC-DNA looping: orientation and distance-dependent loop breaking by the cyclic AMP receptor protein.* Journal of Molecular Biology, 1991. **218**(1): p. 45-54.

25. Lobell, R.B. and R.F. Schleif, *DNA looping and unlooping by AraC protein.* Science (Washington, DC, United States), 1990. **250**(4980): p. 528-32.
26. Kimata, K., et al., *cAMP receptor protein-cAMP plays a crucial role in glucose-lactose diauxie by activating the major glucose transporter gene in coli 10.1073/pnas.94.24.12914.* PNAS, 1997. **94**(24): p. 12914-12919.
27. Saier, M.H., Jr and S. Roseman, *Sugar transport. 2nducer exclusion and regulation of the melibiose, maltose, glycerol, and lactose transport systems by the phosphoenolpyruvate:sugar phosphotransferase system.* J. Biol. Chem., 1976. **251**(21): p. 6606-6615.
28. Pastan, I. and R.L. Perlman, *Repression of b-galactosidase synthesis by glucose in phosphotransferase mutants of Escherichia coli. Repression in the absence of glucose phosphorylation.* Journal of Biological Chemistry, 1969. **244**(21): p. 5836-42.
29. Saier, M.H., Jr and S. Roseman, *Sugar transport. The crr mutation: its effect on repression of enzyme synthesis.* J. Biol. Chem., 1976. **251**(21): p. 6598-6605.
30. Plumbridge, J., *Expression of ptsG, the gene for the major glucose PTS transporter in Escherichia coli, is repressed by Mlc and induced by growth on glucose.* Molecular Microbiology, 1998. **29**(4): p. 1053-1063.
31. Postma, P.W. and S. Roseman, *The bacterial phosphoenolpyruvate:sugar phosphotransferase system.* Biochimica et Biophysica Acta, Reviews on Biomembranes, 1976. **457**(3-4): p. 213-57.
32. Saier, M.H., Jr. and B.U. Feucht, *Coordinate regulation of adenylate cyclase and carbohydrate permeases by the phosphoenolpyruvate:sugar phosphotransferase system in Salmonella typhimurium.* J Biol Chem FIELD Full Journal Title:The Journal of biological chemistry FIELD Publication Date:1975. **250**(17): p. 7078-80. FIELD Reference Number: FIELD Journal Code:2985121R FIELD Call Number:.
33. Simoni, R.D., S. Roseman, and M.H. Saier, Jr., *Sugar transport. Properties of mutant bacteria defective in proteins of the phosphoenolpyruvate: sugar phosphotransferase system.* J Biol Chem FIELD Full Journal Title:The Journal of biological chemistry FIELD Publication Date:1976. **251**(21): p. 6584-97. FIELD Reference Number: FIELD Journal Code:2985121R FIELD Call Number:.
34. Hogema, B.M., et al., *Inducer exclusion by glucose 6-phosphate in Escherichia coli.* Molecular Microbiology, 1998. **28**(4): p. 755-765.
35. Hogema, B.M., et al., *Inducer exclusion in Escherichia coli by non-PTS substrates: the role of the PEP to pyruvate ratio in determining the phosphorylation state of enzyme IIAGlc.* Mol Microbiol FIELD Full Journal Title:Molecular microbiology FIELD Publication Date:1998. **30**(3): p. 487-98. FIELD Reference Number: FIELD Journal Code:8712028 FIELD Call Number:.
36. Fox, D., N. Meadow, and S. Roseman, *Phosphate transfer between acetate kinase and enzyme I of the bacterial phosphotransferase system.* J. Biol. Chem., 1986. **261**(29): p. 13498-13503.
37. Ryu, S., et al., *Effect of the FruR regulator on transcription of the pts operon in Escherichia coli.* Journal of Biological Chemistry, 1995. **270**(6): p. 2489-96.

38. Kimata, K., et al., *Expression of the glucose transporter gene, ptsG, is regulated at the mRNA degradation step in response to glycolytic flux in Escherichia coli.* EMBO Journal, 2001. **20**(13): p. 3587-3595.
39. Morita, T., et al., *Accumulation of Glucose 6-Phosphate or Fructose 6-Phosphate Is Responsible for Destabilization of Glucose Transporter mRNA in Escherichia coli.* Journal of Biological Chemistry, 2003. **278**(18): p. 15608-15614.
40. Peterkofsky, A. and C. Gazdar, *Glucose inhibition of adenylate cyclase in intact cells of Escherichia coli B.* Proceedings of the National Academy of Sciences of the United States of America, 1974. **71**(6): p. 2324-8.
41. Aiba, H., *Autoregulation of the Escherichia coli crp gene: CRP is a transcriptional repressor for its own gene.* Cell (Cambridge, MA, United States), 1983. **32**(1): p. 141-9.
42. Saier, M.H., Jr. and T.M. Ramseier, *The catabolite repressor/activator (Cra) protein of enteric bacteria.* Journal of Bacteriology, 1996. **178**(12): p. 3411-3417.
43. Monedero, V., P.W. Postma, and G. Perez-Martinez, *Suppression of the ptsH mutation in Escherichia coli and Salmonella typhimurium by a DNA fragment from Lactobacillus casei.* J Bacteriol FIELD Full Journal Title:Journal of bacteriology FIELD Publication Date:1998. **180**(19): p. 5247-50. FIELD Reference Number: FIELD Journal Code:2985120R FIELD Call Number:.
44. Kimata, K., et al., *A global repressor (Mlc) is involved in glucose induction of the ptsG gene encoding major glucose transporter in Escherichia coli.* Molecular Microbiology, 1998. **29**(6): p. 1509-1519.
45. Bruckner, R. and F. Titgemeyer, *Carbon catabolite repression in bacteria: choice of the carbon source and autoregulatory limitation of sugar utilization.* FEMS Microbiology Letters, 2002. **209**(2): p. 141-148.
46. Hogema, B.M., et al., *Autoregulation of lactose uptake through the LacY permease by enzyme IIAGlc of the PTS in Escherichia coli K-12.* Molecular Microbiology, 1999. **31**(6): p. 1825-1833.
47. Ackerman, R.S., N.R. Cozzarelli, and W. Epstein, *Accumulation of Toxic Concentrations of Methylglyoxal by Wild-Type Escherichia coli K-12.* J. Bacteriol., 1974. **119**(2): p. 357-362.
48. Dien, B.S., N.N. Nichols, and R.J. Bothast, *Fermentation of sugar mixtures using Escherichia coli catabolite repression mutants engineered for production of L-lactic acid.* Journal of Industrial Microbiology & Biotechnology, 2002. **29**(5): p. 221-227.
49. Lindsay, S.E., R.J. Bothast, and L.O. Ingram, *Improved strains of recombinant Escherichia coli for ethanol production from sugar mixtures.* Applied Microbiology and Biotechnology, 1995. **43**(1): p. 70-5.
50. Snoep, J.L., et al., *Involvement of pyruvate dehydrogenase in product formation in pyruvate-limited anaerobic chemostat cultures of Enterococcus faecalis NCTC 775.* Arch Microbiol FIELD Full Journal Title:Archives of microbiology FIELD Publication Date:1990. **154**(1): p. 50-5. FIELD Reference Number: FIELD Journal Code:0410427 FIELD Call Number:.
51. Hernandez-Montalvo, V., et al., *Characterization of sugar mixtures utilization by an Escherichia coli mutant devoid of the phosphotransferase system.* Applied Microbiology and Biotechnology, 2001. **57**(1-2): p. 186-191.

Metabolic Flux Analysis of *Escherichia coli* under Anaerobic Conditions and Design of ^{13}C labeling Experiments

4.1 Introduction

Metabolic flux analysis (MFA) is an important tool in metabolic engineering[1]. Flux measurements and the comparisons of fluxes between different phenotypes can provide insights toward selection of appropriate metabolic engineering targets [2]. The classical approach of analyzing intracellular carbon fluxes is called conventional flux analysis (C-MFA). It is based on metabolite balances around intracellular metabolites with the measurements of extracellular fluxes as constraints for flux calculation. Pseudo-steady state i.e. the rate of change of intracellular metabolite concentration is small compared to fluxes in and out of the metabolite, is assumed. Frequently, the lack of enough measurements requires assumptions about redox (NADH/NADPH) or energy balances. However, incomplete pathway knowledge with NADH/NADPH or ATP flux balances, which is very common as they are involved in large number of reactions, can lead to incorrect flux estimation [3]. The development of carbon labeling (^{13}C) experimentation provided researchers with the extra measurements so to avoid assumptions about the redox and energy balances and therefore, enhances the reliability of the flux estimates. Although, ^{13}C MFA has applied extensively for *Escherichia coli* growing under aerobic conditions[4-7], there are a few literature reports pertaining to the

^{13}C flux analysis under anaerobic conditions. Most of the flux analyses under anaerobic conditions have been done by conventional MFA with a simplified metabolic model. Szyperski estimated metabolic flux ratios at important metabolic branch points under anaerobic conditions[8]. The NMR data were used calculate the bond integrity of precursor molecules which were in turn used to estimate flux ratios. Using the same NMR data, Schmidt *et al* [3] carried out the first comprehensive ^{13}C MFA where metabolic balance is used in conjunction with constrains from labeling measurements. However, extracellular measurements used in flux estimation criteria were obtained from different experiment [9]. Using an approach similar to those by Szyperski[8], Sauer *et al* [10] reported ^{13}C 'METAFOR' analysis under anaerobic conditions to calculate flux ratios at few nodes. These flux ratios are the additional equations to be used in conjunction with mass balances to solve system wide fluxes. Since flux ratios are based on a simplified network topology, thus they are not as complete as the comprehensive ^{13}C flux analysis.

In this work, we are going to present the C-MFA and also comprehensive ^{13}C MFA of *E. coli* under anaerobic conditions. We found that the extracellular measurements played an important role even in ^{13}C flux analysis. Flux identifiability i.e. whether flux is indeed identifiable or not, is substantially altered by varying the labeling state of substrate fed to the system. Hence, identifiability analysis was carried out for the optimal design of ^{13}C experiment when some of the extracellular measurements are not used in flux analysis.

4.2 Material and Methods

Bacterial strains and cultural conditions

The wild type strain of *Escherichia coli* K-12 (W3110) was used in this work. The strain was a gift from Dr. L.O Ingram to Dr. Ramon Gonzalez. The strain from glycerol stocks at -80˚C, were streaked onto LB plates and incubated for 12 hours at 37˚C. The colonies from the LB plate were used to inoculate two 50 mL tubes completely filled with minimal media supplemented with 10g/L of glucose. The tubes were incubated at 37˚C until an OD_{550} of 0.6 was reached. This actively growing pre-culture was centrifuged at 5000g for 15 min at +4˚C and the pellet was reconstituted in the media and the appropriate volume of the reconstituted pellet was used to inoculate 900mL of medium in the fermenter, with the target starting OD_{550} of 0.05.

The batch fermentation was conducted at 37^{o}C in 1L bioreactor (Bioflow 110, New Brunswick Scientific, Edison, New Jersey) fitted with offline CO_2 gas analyzer. The minimal media with the initial glucose concentration of 10g/L was used to grow the cells. The pH was controlled at 6.8 by adding 4M KOH and stirrer speed was set to 200 rpm. The initial working volume was 900 mL. Anaerobic condition was maintained by flushing the head space with high purity nitrogen.

Two labeling experiments were carried out: one with 10% U-^{13}C glucose(Cambridge Isotope Limited, Andover, MA) and 90% naturally labeled glucose and another with 10% U-^{13}C glucose, 25% 1-^{13}C glucose(Cambridge Isotope Limited, Andover, MA) and 65% naturally labeled glucose.

Analytical procedures

Dry cell weight was monitored by measuring optical density (OD_{550}) (Genesys 20, Thermospectronic, and Madison, Wisconsin)(1OD=0.36g DW/L). Glucose and

fermentation product were measured on Waters (Milford, MA) HPLC system with a 410 refractive index (RI) detector. The Aminex column (HPX-87H, Bio-Rad, Hercules, CA, USA) was maintained at 42˚C and 5mM H_2SO_4 was used as the mobile phase at flow rate of 0.3mL/min. Carbon dioxide in the off gas was monitored by gas analyzer of the fermenter.

Sample Preparation for 2D NMR measurement

For isotopomer analysis by NMR, 750 mL of biomass was harvested at OD_{550} of 0.6 and the cell growth was quenched by keeping the cells on ice water. The cells were first centrifuged at 5000g for 15 min at +4˚C. The pellet were washed with 0.9% saline water and centrifuged again for 15min at 5000 g. About 80mg of biomass(estimated dry weight) was transferred in four hydrolysis tubes (Pierce Endogen, Rockford, IL), to which 6 N hydrochloric acid(Pierce Endogen, Rockford, IL) was added in the 1 mL of HCl:4 mg of biomass. The hydrolysis tube was evacuated, flushed with nitrogen to remove residual oxygen, and reevacuated. The hydrolysis was performed at 110°C for 12 hrs. The acid in the hydrolyzates was evaporated in a Rapidvap evaporator (Labconco, Kansas City, MO). The residue was reconstituted in 2 ml of deionized water, lyophilized for 72 h, and finally the dissolved in 500 μL of D_2O(Sigma, St. Louis) in an NMR tube. The pH of the NMR sample was adjusted to 1 using DCl (Sigma, St. Louis).

NMR experiment

Two-dimensional [^{13}C, ^{1}H] HSQC NMR spectra were collected on a Bruker Avance DRX 500 MHz spectrometer (Bruker Instruments, Billerica, MA) at 298 K. The reference to 0 ppm was set using the methyl signal of dimethylsilapentanesulfonate (Sigma, St. Louis) as an internal standard. The resonance frequency of ^{13}C and ^{1}H were

125.7 MHz and 499.9 MHz, respectively. The spectral width was 5,482.26 Hz along the ^{1}H (F2) dimension and 5,028.05 Hz along the ^{13}C (F1) dimension. Peak aliasing was used in order to minimize the sweep width along the F1 dimension. The number of complex data points was 1,024 (^{1}H) x 900 (^{13}C). A modification of the INEPT (insensitive nuclei enhanced by polarization transfer) pulse sequence was used for acquiring HSQC spectra. The number of scans was generally set to 16.

The software Xwinnmr (Bruker Instruments, Billerica, MA) was used to acquire all spectra, and the software NMRView[11] was used to quantify nonoverlapping multiplets on the HSQC spectrum. Overlapping multiplets (α-amino acid), which could not be processed with NMRView, were quantified by a peak deconvulution software based on spectral processing algorithm proposed by Van Winden *et al* [12]. The standard deviations associated with the NMR intensity measurements were estimated from the noise to peak intensity ratio with minimum set to 1%.

Metabolic flux calculation

The pools of intracellular metabolite were assumed to be in isotopic steady state. Additionally, pseudo steady state was assumed for flux and metabolite i.e. intracellular metabolite concentration and fluxes do not vary during the experiment. Since initial OD was less than 10% the final OD, the dilution effect to HSQC labeling measurement due to initial unlabeled biomass was found to be negligible (Appendix). The metabolic network model used for the flux estimation is shown in and also included in Appendix (). The model consists of reaction in central carbon metabolism namely Embden-Meyerhof-Parnas pathway (EMP) and Pentose phosphate pathways (PP), TCA cycle, glyoxylate

shunt and reversible anaplerotic fluxes from phosphoenolpyruvate (PEP) to oxaloacetate (OAA). Transport reactions for the fermentation products were also included. The carbon fate of precursor in proteinogenic amino acids was constructed based on that described by Szyperski [8]with exception of serine and glycine pathway. The synthesis of serine from 3-phospho-glycerate (3PG) was included. One carbon metabolism of serine to glycine was also included in the model. Additionally, fermentation reactions leading to acetate and ethanol from acetyl-CoA were combined as it leads to similar carbon rearrangement. The triose pool of dihydroxy-acetone phosphate(DHAP) and 3PG were also lumped. The high exchange between ribose 5 phosphate (R5P) and xylose 5-phosphate (X5P) was also observed. Hence, pentose pools were also lumped into R5P. The reactions leading to oxidative pentose pathway (Zwf) and TCA were considered irreversible with no negative flux allowed through these reactions. The reaction through Pgi, Eno in EMP pathway and Tkt, Tkt2, Tal in pentose phosphate pathway were considered reversible. The flux though Pfl was also assumed to be reversible. Since succinate is a symmetric molecule, the scrambling reaction was also included in the model. The forward (V_1) and backward flux (V_{-1}) associated with each reversible reaction step were transformed into net flux (V_{net}) and extent of reversibility (r)

$$V_{net} = V_1 - V_{-1}$$

$$r = \frac{\min(V_1, V_{-1})}{\max(V_1, V_{-1})}$$

In order to avoid numerical problems, the extents of reversibility were constrained between 0 and 0.99.

The biomass fluxes of precursor molecules were estimated from the literature data using biomass yield (Appendix A). In addition to extracellular flux of glucose, lactate, succinate, the flux through TCA (Suc), oxidative pentose pathway (Zwf), glyoxalate shunt (Gos) and biomass fluxes were chosen as free fluxes. The extents of reversibility were also considered as free parameters.

The flux evaluation program NMR2Flux[13] developed for soybean network was modified to generic flux evaluation program in which user can supply metabolic network as inputfile. The NMR2Flux[13] estimates flux by minimizing the difference between simulated and experiment NMR intensities. First, free fluxes are guessed which are then used to calculate all intracellular fluxes using stoichiometric balance. The calculated fluxes are then used to estimate the labeling pattern for the intracellular metabolite by the solving cummomer balances of all the metabolite in the network. The simulated NMR intensities for the proteinogenic amino acids are estimated from the labeling pattern of the precursor metabolite for each amino acid. The chi-square (χ^2) is calculated from the difference between simulated and experimental intensities. The extracellular flux measurements of acetate-ethanol and formate have also been added to the chi-square (χ^2).

$$\chi^2 = \frac{(I_{sim} - I_{\exp})^2}{N_{\exp}^{\ 2}} + \frac{(F_{mes} - F_{sim})^2}{N^2}$$

The set of fluxes which gives minimum χ^2 is taken as real fluxes. To verify the global error minimum, multiple simulations were carried out from different starting points. The statistical error analysis was performed by using Monte Carlo simulation approach in which synthetic NMR intensities were used as surrogate for experimental

data. Finally, the 100 set of flux distributions obtained by Monte Carlo simulation were used to calculate standard deviations (SD) of the fluxes.

Identifiability Analysis

To find the effect of different kind of labeled substrate on the standard deviation of the fluxes, identifiability analysis was carried out as described previously [14]. The standard deviation of the metabolic fluxes was used as a measure of the quality of the labeling experiment and to compare how an experiment with certain choice of substrates performs in relation to another experiment with a different choice of substrate. In this work, the reciprocal of geometric mean of standard deviation of the fluxes is defined as the information content (IC) and was used as the objective criterion of the optimal experiment.

$$IC = \frac{1}{\left(\prod_{i=1}^{n} SD_i\right)^{1/n}}$$

The computational approach of this work is along the lines of Mollney[].If F, X and I denote vectors representing the fluxes, isotopomer distribution of metabolite in the metabolic network and NMR intensities of the amino acids respectively. The relationship between F and X :

$$V = f(X)$$

is nonlinear whereas the relationship between X and I is linear, and can be expressed by matrix manipulation

$$I = M.X$$

Additionally, all fluxes in F are not linearly independent but are constrained by stoichiometric balance. Hence, an independent subset of fluxes (free fluxes) can be used

to estimate all the fluxes. If P represents free fluxes parameters, the relationship between free fluxes parameters and dependent fluxes is

$$F = G.P$$

Together, this relationship can be consolidated into a nonlinear relationship between the flux parameters and the NMR intensities.

$$I = h(P)$$

The NMR2Flux developed previously in our group evaluates flux parameters iteratively from the labeling data.

The covariance matrix of the parameters ($Cov(P)$) can be constructed as

$$Cov(\underline{P})_{ij} = \sum_{m=1}^{m=N_m} \frac{dI_m}{dP_i} . \sum_{m=1}^{m=N_m} \frac{dI_m}{dP_j}$$

Where

N_m = number of NMR measurement.

The differentiation $\left(\frac{dI_m}{dP_i} \right)$ can be done analytically as well as numerically. In this work, numerical differentiation was employed.

The D-criterion, which measures the volume of the confidence ellipsoid of the evaluated flux parameters, is equal to the determinant covariance matrix of flux parameters.

$$D = \det(\underline{\underline{Cov(\underline{P})}})$$

For n number of flux parameters, the n^{th} root of D ($\sqrt[n]{D}$) is the geometric mean of the standard deviation. Hence, we have

$$IC = \frac{1}{\sqrt[n]{D}}$$

All computations of IC are reported with respect to a reference experiment with 10% U-^{13}C glucose experiment as only labeled substrate. The various combinations of U-^{13}C glucose, 1-^{13}C glucose and unlabeled glucose were examined for their ability to provide an improved labeling data set. The values of fluxes at which identifiability was determined were the ones evaluated with reference experiment with 10% U-^{13}C glucose.

4.3 Results and Discussion

Physiological analysis

The anaerobic batch fermentation of wild type *E. coli* was carried out in a 1L fermenter with glucose as the only carbon source. The experimentally determined growth parameters are summarized in Table 4.1. The maximum specific growth rate was found to be 0.60 hr^{-1} which is much higher than previously reported value of 0.3 hr^{-1} for various wild type strains of *E. coli* under anaerobic conditions[10]. Biomass yield was found to be 0.1276±0.001 g g^{-1} of glucose, which is again higher than the previously reported value (0.065 to 0.0975) for various strains of *E. coli* [10] and which might explain the higher growth rate observed. The succinate, lactate, ethanol, acetate and formate were the major fermentation products. The 93±7% of carbon from glucose consumed was recovered either as the fermentation product or as biomass. The yield of ethanol and acetate were high compared to other fermentation products. This might be explained by the fact that in the absence of external electron acceptor under anaerobic condition, redox balance plays very important role in influencing intracellular fluxes. The production of

equimolar amount of ethanol and acetate from glucose is redox balanced. Moreover, it is the most efficient anaerobic mode which produces three mole of ATP for one mole of glucose consumed. However, they are not produced in exactly equal amount because biomass synthesis and other pathways alter the redox balance.

Conventional Metabolic flux Analysis

To gain deeper insight into the intracellular carbon flux distribution, metabolic flux analysis was performed. First, the conventional metabolic flux analysis(C-MFA) was done in which only extracellular measurements are used. The metabolic network model had to be simplified for C-MFA as it is, otherwise, underdetermined. The TCA cycle is reported to be branched under anaerobic conditions and the glyoxalate shunt pathway has also found to be repressed in wild type with the growth on glucose [7]. Hence, TCA and glyoxalate shunt pathway were assumed to be inactive for C-MFA. Since, it is not possible to estimate reversibility in C-MFA, hence the reversibility of reactions was not considered. With these simplifications, the metabolic model is exactly determined. The condition number of the stoichiometric matrix consisting mass balance of intracellular metabolite was found to be 8(Appendix A). It measure the how accurate is the flux calculated from the stoichiometric model with respect to variation in measurement. For C-MFA with extracellular measurement, the condition number should be between 1 and 100 [].

The extracellular measurements of the fermentation product and biosynthetic fluxes were used to calculate the intracellular fluxes. The Monte Carlo simulation

approach was used to estimate the standard deviation of intracellular fluxes. shows metabolic flux map of *E. coli* with standard deviations. The fluxes are reported with relative to 100 moles of glucose.

The flux through pyruvate formate lyase (Pfl) catalyzing conversion of pyruvate to acetyl-CoA and formate was found to be 147±10 suggesting very high activity under anaerobic conditions. The combined flux of subsequent conversion of acetyl-CoA to ethanol and acetyl-CoA to acetate flux was found to be 136±10. The production of equimolar amount of ethanol and acetate from glucose is the most efficient anaerobic mode producing three molecules of ATP per glucose molecule fermented and is redox balanced. Thus, it explains very high flux through Pfl. The flux through lactate (Ldh) and succinate (Succ) were found to be 11.44±7.7 and 5.54±1.42. The pathway leading to succinate is not redox balanced. On the other hand, although the production of lactate from glucose is redox balanced, it produces two moles of ATPs per mole of glucose consumed. Hence, it is less efficient than acetate-ethanol pathway.

The flux through oxidative branch of pentose phosphate pathway (Zwf) was found to be 38±37 suggesting that pentose pathway may be active under anaerobic condition which is consistent with previous finding by METAFor[10]. Using ^{13}C flux analysis, Schmidt *et al* [3] found the flux through ZWF to be 77 % under anaerobic condition.

The flux through Zwf was found to be associated with large standard deviation. The solution matrix was constructed to find the sensitivity of calculated fluxes with measured fluxes (Appendix A). The flux through Zwf was found to be sensitive to the error in biomass fluxes and HPLC measurements of glucose, lactate, and acetate-ethanol.

The large SD observed in the flux through Zwf is due to the large error in the acetate and lactate extracellular flux measurements.

^{13}C Labeling Metabolic flux Analysis

The ^{13}C labeling data provide additional measurement data which can be used to verify the results of C-MFA as well as the metabolic network model. It can also be used to increase the resolution of the metabolic fluxes. The ^{13}C labeling experiment was carried out with 10% U-^{13}C glucose and 90% naturally labeled glucose (referred as U-^{13}C experiment). The exponentially growing cells were harvested and hydrolyzed with 6N HCl. The 2D HSQC spectrum was obtained (figure1) to obtain ^{13}C labeling pattern of proteinogenic amino acids. The effect of dilution due to the initial biomass on HSQC measurement was found to be negligible and, hence, was neglected.

TCA and glyoxalate shunt reaction were added to the stoichiometric model. The reversibility's of reactions was also included in the model. There are 37 net reactions rates and 15 reversible reactions i.e. total of 52 fluxes to be estimated. Assuming pseudo steady state, 20 intracellular metabolite pools contribute to 20 linear constraints. Hence, the model has 32(52- 20) parameters including 15 reversible parameters. The 12 biomass fluxes of precursor molecules were estimated from the literature data using biomass yield (Appendix A). The extracellular flux of glucose, lactate and succinate were calculated from the HPLC measurements. The extracellular fluxes of formate and the combined flux of acetate-ethanol were included in the chi-square (χ^2) criteria.

NMR2Flux was used to estimate metabolic fluxes which best account for NMR measurements and HPLC measurements. In general, there was good fit between the simulated and experimental measurement. The total chi-square was 650 with average

difference between simulated and experimental intensities to be 0.019. Out of 105 NMR measurements, 27 measurements from asp-α, ile- α, phe- α, leu- α, tyr-β and tyr-δ contributed to 60% of the total chi-square. Most of these peaks were analyzed by spectral deconvulution. Therefore, their high contribution to chi-square is most likely due to the low error assumed for these peaks rather than the wrong metabolic model.

The metabolic flux map estimated by NMR2Flux is shown in Figure 4.2. The flux values are reported with relative to 100 moles of glucose consumed. The MFA results indicate no flux through SUC indicating TCA is not complete under anaerobic conditions, in agreement with previous studies[3, 8, 10]. Hence, TCA cycle is branched which operates mainly to fulfill demand for precursors for biomass synthesis.

The pentose pathway provides cells with CO_2 and NADPH needed for growth. It also produces precursors like R5P for biomass synthesis but their metabolic demand is small due to low biomass flux under anaerobic conditions. The relative flux through oxidative part of pentose pathway (Zwf) was found to be 40±30. Using semi-quantitative NMR analysis, Szyperski *et al* [8] estimated that 20 to 30% of glucose converted to PEP via the pentose pathway in *E coli* B. Additionally, they found that around less than 20% of R5P originates from G6P via ZWF. However, it is not possible compare these results directly with net fluxes obtained by C-MFA or ^{13}C MFA because rapid equilibration of pentose pool in addition to rapid exchange via transketolase and/or transaldose can lead to similar carbon labeling pattern for various intracellular flux distribution[]. Using same NMR data, Schmidt *et al* [3] found flux through Zwf to be 77% from the comprehensive ^{13}C flux analysis. However, the extracellular flux measurement for the fermentation

products were taken from the literature[9]. Additionally, no statistical analysis was carried out which makes interpretation of the result difficult.

The glyoxalate shunt is often assumed to be inactive in glucose grown cultures as it is subjected to catabolite repression by glucose. However, recently glyoxalate shunt was found to be active in wild-type *E. coli* in a glucose-limited chemostat [15]and under glucose-excess batch conditions in the case of *Pgi* (phosphoglucose isomerase) mutant [15, 16]. Hence, these finding opens up the question whether glyoxalate shunt is indeed inactive in glucose grown culture [17]. Our MFA indicate the flux through glyoaxate shunt (Gos) to be 2.3±2.5. Additionally, χ^2 didn't increase (650 to 729) if both TCA and GOS were not included in the model. Hence, it can be concluded that GOS is not active the experimental conditions.

Redox Balance and ATPs

The estimated flux distribution was used to calculate total redox balance (NADH+NADPH) which is required for biomass formation and produced by oxidative pentose phosphate pathway (ZWF), isocitrate dehydrogenase (CS)), glyceraldehydes 3-phosphate dehydrogenase(ENO), acetaldehyde dehydrogenase(AC) and lactate dehydrogenase (LDH). According to our calculation, there was a deficiency of redox equivalents (NADH+NADPH). However, flux estimated with 1-^{13}C glucose and U-^{13}C glucose without acetate/ethanol measurement (see statistical analysis), redox balance was observed.

The estimated flux distributions also allow the calculation the balance of NADPH. According to our calculation, NADPH imbalance is observed probably due to the interconversion between NADH and NADPH catalyzed by the transdehydrogenase [16].

The transdehydrogenase flux converting NADH to NADPH is found to be 42±2 for wild type *E. coli.*

Unlike in the case of aerobic conditions, it is possible to estimate metabolic production of ATP from the flux without assuming P/O ratio. The total ATP production in central carbon metabolism was found to be 164±0.5. Several cellular processes require the consumption of ATP to ensure the proper functioning of cells without net generation of cell biomass and they are called maintenance energy. ATP maintenance was found to be 120±0.55.

Statistical and Identifiability Analysis

The standard deviation associated with pentose flux was found to be [] and [] in ^{13}C-MFA with 10% U-^{13}C glucose and C-MFA respectively. Our ^{13}C flux methodology imposes irreversibility constraints on Zwf flux, while C-MFA analysis no such constrained was imposed (see our previous paper []). Hence, to find the effect of such constrains on SD of Zwf, irreversibility constrain was imposed on in C-MFA analysis. The flux map is shown in figure 1.2. Although, the SD of ZWF flux decreased from 34 to 22, it is still higher than that those estimated in U-^{13}C experiment. Hence, it can be concluded that the labeling measurements in U-^{13}C experiment do provide additional information about ZwF flux, however, it is insufficient to accurately determine Zwf flux.

In comprehensive ^{13}C-MFA, the NMR measurements are used along with extracellular measurements in order estimate intracellular fluxes. As such, the labeling fluxes are dependent on intracellular fluxes only and do not contain information about extracellular fluxes. However, since intracellular fluxes are related to extracellular fluxes by stoichiometric constrains, the accurate determination of intracellular flux would result

in accurate determination of extracellular fluxes. To test this hypothesis whether extracellular fluxes can be estimated from labeling measurement alone, formate and ethanol-acetate measurements were not included in the chi-square criteria. Figure 4.5 lists the result of ^{13}C MFA without these extracellular measurements. The acetate-ethanol flux was estimated to be 137±10 which is similar to that estimated when extracellular measurement was included in the chi-square. The measured flux value for acetate-ethanol flux was 136±8. Hence, acetate-ethanol combined flux can be estimated more accurately from the labeling alone.

On the other hand, compared to the case when extracellular measurements were not used in chi-square, the formate, Fdhf and CO_2 fluxes were found to be associated with much larger SDs. However, the SDs of combined flux of Fdhf and formate was found to be much lower. Hence, the conversion of formate to CO_2 through Fdhf is unidentifiable under the experimental condition. However, the flux could be estimated if the flux through Fdhf affects the labeling pattern of CO_2 and thereby affecting the labeling pattern of amino acids originating from the oxaloacetate (the 4^{th} carbon of oxaloacetate come from CO_2). Hence, in principle, formate flux can be estimated from the labeling measurement if the enrichment of CO_2 and formate are different. In ^{13}C labeling experiment with U-^{13}C glucose and naturally labeled glucose as the only carbon source, the enrichment of all carbon is same (11% in the present case).Thus, formate and Fdhf flux is unidentifiable in U-^{13}C experiment.

Large standard deviation associated with Zwf flux suggests that the labeling measurement do not provide enough information to estimate fluxes through Zwf. This is in agreement with previous finding by Dauner *et al*[18] who found large confidence

interval with Zwf flux in their MFA with U-^{13}C glucose. Nevertheless, Schmidt *et al*[19] were able to accurately estimate Zwf flux using a mixture of 1-^{13}C glucose and U-^{13}C glucose. Hence, identifiability analysis was carried out for the optimal labeling mixture. Various combinations of U-^{13}C glucose and 1-^{13}C glucose were examined for their effect on flux identifiability at the flux values estimated from the ^{13}C labeling experiment. Figure 4.7 depicts the statistical quality for various combinations relative to the reference experiment. The information content (IC) which indicate the statistical quality of the experiment, is reported to the reference experiment with 10% U- ^{13}C labeled glucose. In computation of IC , the statistical quality of all the flux parameters was taken into the account. The maximum improvement in information content (IC) of 120% was observed with the combination of 5% U-^{13}C glucose and 95% 1-^{13}C glucose. There is a huge improvement in statistical quality of Zwf flux when 1-^{13}C glucose is used with U-^{13}C glucose as labeled substrate (shown in Figure 4.8).

Identifiability analysis uses linear statistical method to estimate standard deviation of fluxes, which is consecutively used to calculate IC . Linearized methods are based on a linear approximation of the nonlinear relation around the estimated flux; it may not give correct estimate of standard deviation (SD) of fluxes. Additionally, it doesn't account for multiplicity of solution or non-ideal behavior (sub-optimal) of optimization techniques employed in flux estimation [3]. Furthermore, it has been found to give overestimate of the SDs of the fluxes in case of PYR-PEP-OAA triangle [18]. Hence, it may not give correct estimate of SD of fluxes and thus identifiability analysis based on linearized methods may not be the most reliable way to design labeling experiment.

Since Monte Carlo simulations (MC) were used for the statistical analysis of flux, MC was used to verify the results of identifiability analysis. This was done by estimating SDs of flux distributions from the simulated NMR intensities for various input labeled substrate. These simulated NMR intensities were obtained by simulating the labeling experiment for the fluxes estimated in the U-^{13}C experiment. Table 4.3 lists the results of the simulation. The statistical quality of Zwf shows improvement with the use of 1-^{13}C glucose. The substrate combination is the 25% of 1-^{13}C glucose and 10% of U-^{13}C glucose was chosen as optimal mixture as it estimates Zwf flux accurately.

The labeling experiment was carried out with the optimal mixture of 10% U ^{l3}C glucose, 25 % 1-^{13}C glucose and 65% naturally labeled glucose (referred as 1-^{13}C experiment). MFA analysis was done without extracellular measurement in the chi-square. The flux map is shown in Figure 4.6. In agreement with *in silico* analysis by MC, the statistical quality of Zwf showed an improvement of 18 times with SDs decreased to 0.72 from 14. This resulted in more accurate determination of glycolytic and pentose fluxes. All fluxes in pentose pathway including reversibility showed improvement with 1-labeled experiment for example, tkf showed an improvement of 12 times. Glyoxalate shunt pathway (*GOS*) flux also showed an improvement of 4 times. However, it has poor resolution compared to U-^{13}C experiment with extracellular measurements. The combined extracellular flux of acetate and ethanol was also estimated accurately. It is not possible to estimate extracellular fluxes of ethanol and acetate independently as they have same carbon rearrangement.

Formate and Fdhf flux were, however, associated with large SD when extracellular measurements are not used in the chi-square. As discussed previously, the Fdhf flux

could be estimated if the enrichment of formate and CO_2 differs. To probe why formate flux was associated large SD in 1-^{13}C experiment; linearized method was used to estimate SD of formate for various flux values of Zwf, TCA and Ppc which are the main reaction producing CO_2. It can be seen that TCA and Ppc do not have much effect on the SD of formate flux. On the other hand, Zwf flux has higher flux effect on the SD of formate flux. The flux through Zwf affects the fraction of CO_2 from 1st carbon of glucose. Hence, higher Zwf flux results higher enrichment of CO_2.However, under the experimental condition, small flux through Zwf results in similar enrichment of formate and CO_2.

The generic nature of our flux evaluation methodology allows us to easily modify metabolic network model used for MFA. Hence, various modifications to the original network models were tested for their ability to explain the experiment measurements. The MFA analysis without TCA and glyoxalate shunt didn't affects the chi-square in both U-^{13}C and 1-^{13}C experiments. Additionally, the addition of glyoxalate shunt resulted in change in fluxes of mal, frd in both U-^{13}C and 1-^{13}C experiment. However, there was not much effect on other fluxes distribution. The MFA analysis with ED and MEZ resulted in comparably higher decrease in chi-square in both U-^{13}C and 1-^{13}C experiment. The addition of these two reactions led to better fit of gly-α and phe- α in 1-^{13}C experiment and asp- α, ile- α for U-^{13}C experiment. ED pathway was found to be unidentifiable in U-^{13}C experiment. The flux through ED pathway was found to be 1.3 with low SD deviation of 1.5. Hence, ED pathway is not active under anaerobic growth on glucose. The MEZ flux was, on the other hand, was found to be unidentifiable both in U-^{13}C and 1-^{13}C experiment. Furthermore, the large variation in MEZ flux was correlated with large variation in MAL, FALR fluxes because of stoichiometric constraints.

4.4 Conclusion

C-MFA and ^{13}C-MFA of wild type *E. coli* was carried out under anaerobic condition. Both C-MFA and ^{13}C-MFA indicated high activity of pyruvate formate lyase (Pfl) under anaerobic conditions. The flux through oxidative pentose pathway (Zwf) was found to be associated large SDs in both C-MFA and ^{13}C MFA experiment with 10% U-^{13}C experiment as the labeling substrate. It was also found that large error associated with Zwf flux is due to large error in extracellular measurements. It was further concluded, labeling measurement do not provide enough information to estimate Zwf flux. Hence, accurate estimation of Zwf flux is dependent on accuracy of extracellular measurement in ^{13}C MFA with U-^{13}C glucose as the labeling substrate. However, we showed that it possible to estimate Zwf flux even without using some of the extracellular measurement in flux analysis. In fact, it is possible to estimate those flux from the labeling measurements.

Table 4.1: Growth parameters of W3110 *E. coli* in glucose grown anaerobic culture.

Biomass Yield (g g-1)	0.13
Glucose uptake rate(mmol g-1 h-1)	8.60
C balance	0.93
Ethanol Yield(mol / mol)	0.62
Acetate Yield(mol / mol)	0.71
LactateYield(mol /mol)	0.12
Succinate Yield(mol / mol)	0.55
Formate Yield(mol / mol)	1.36

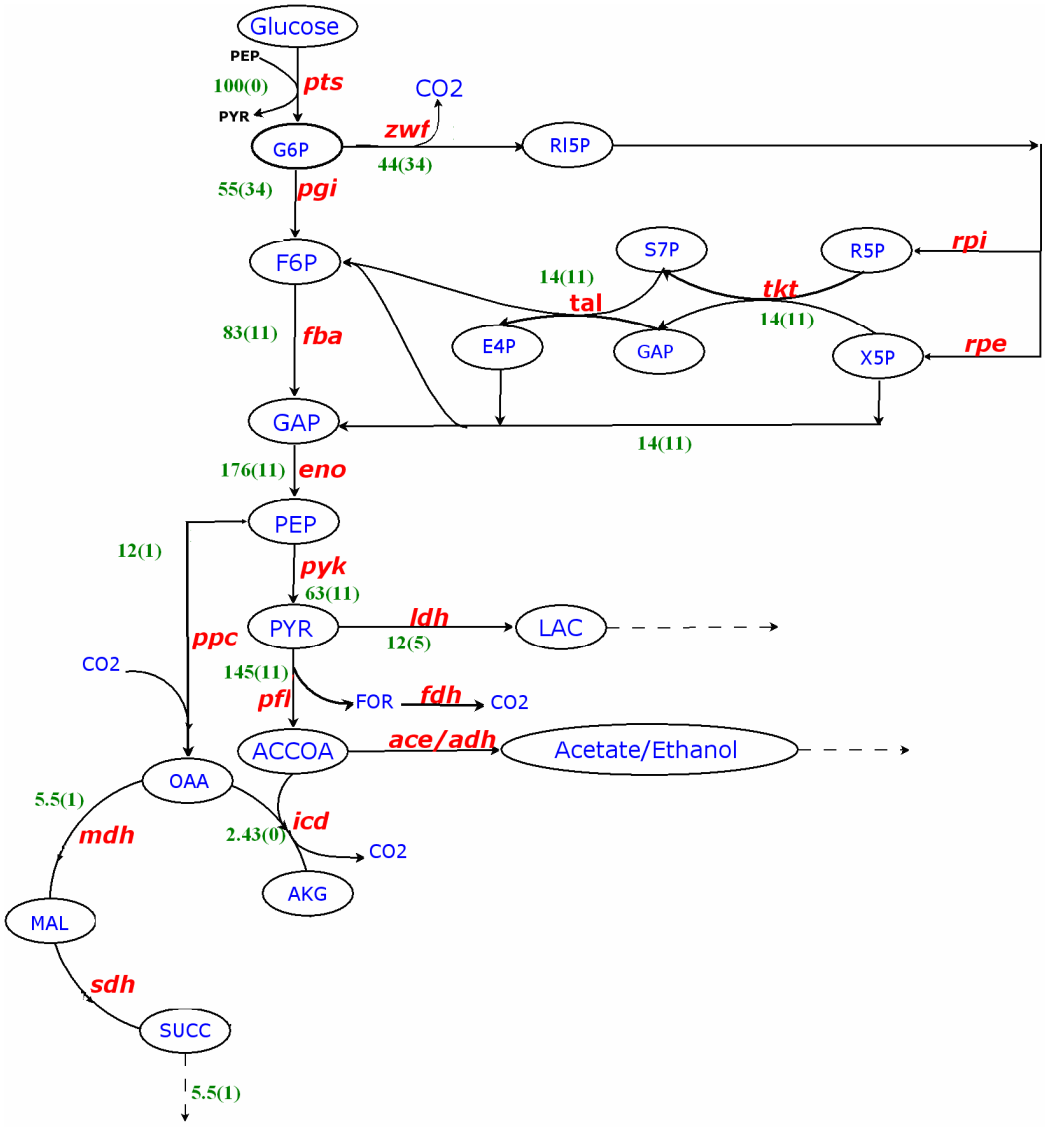

Figure 4.1 : Metabolic network representing central carbon metabolism in *E. coli*. This figure also represents the metabolic flux map of W3110 *E. coli* under anaerobic conditions obtained via conventional flux analysis. The flux values are reported relative to 100 moles of glucose consumed.

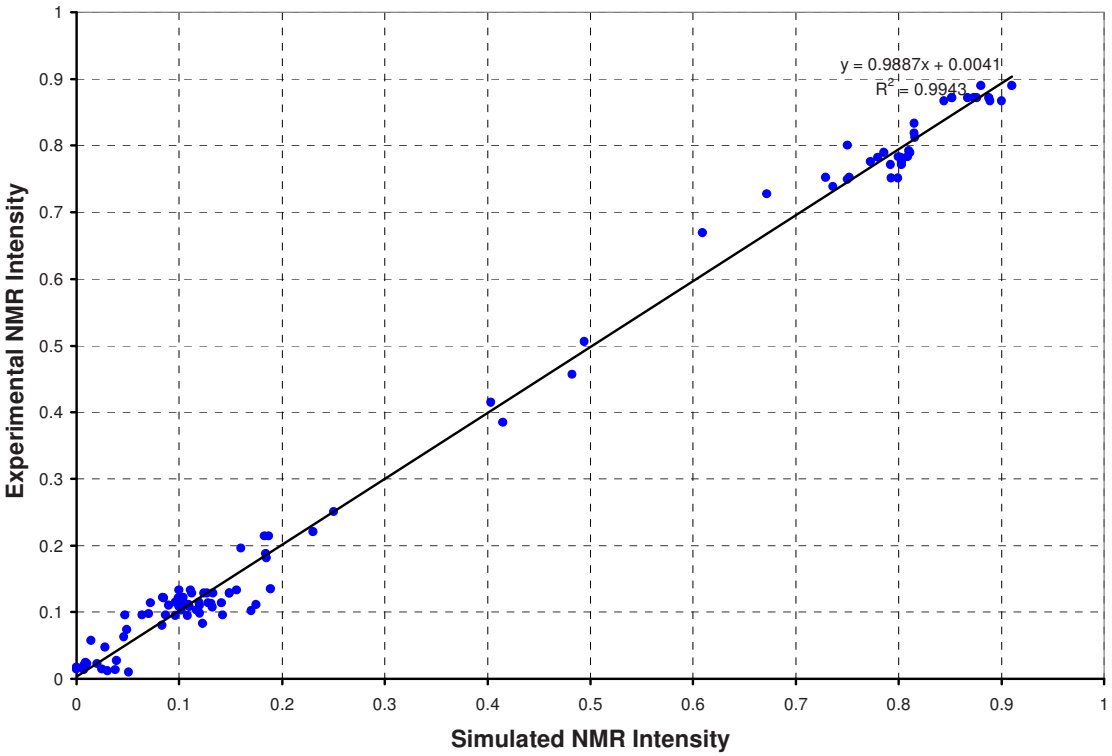

Figure 4.2: Comparison of experimental and simulated NMR intensities of amino acids from *E. coli* biomass hydrolysate of ^{13}C labeling experiment with 10%U^{13}-C glucose and 90% ^{12}C glucose as the carbon source.

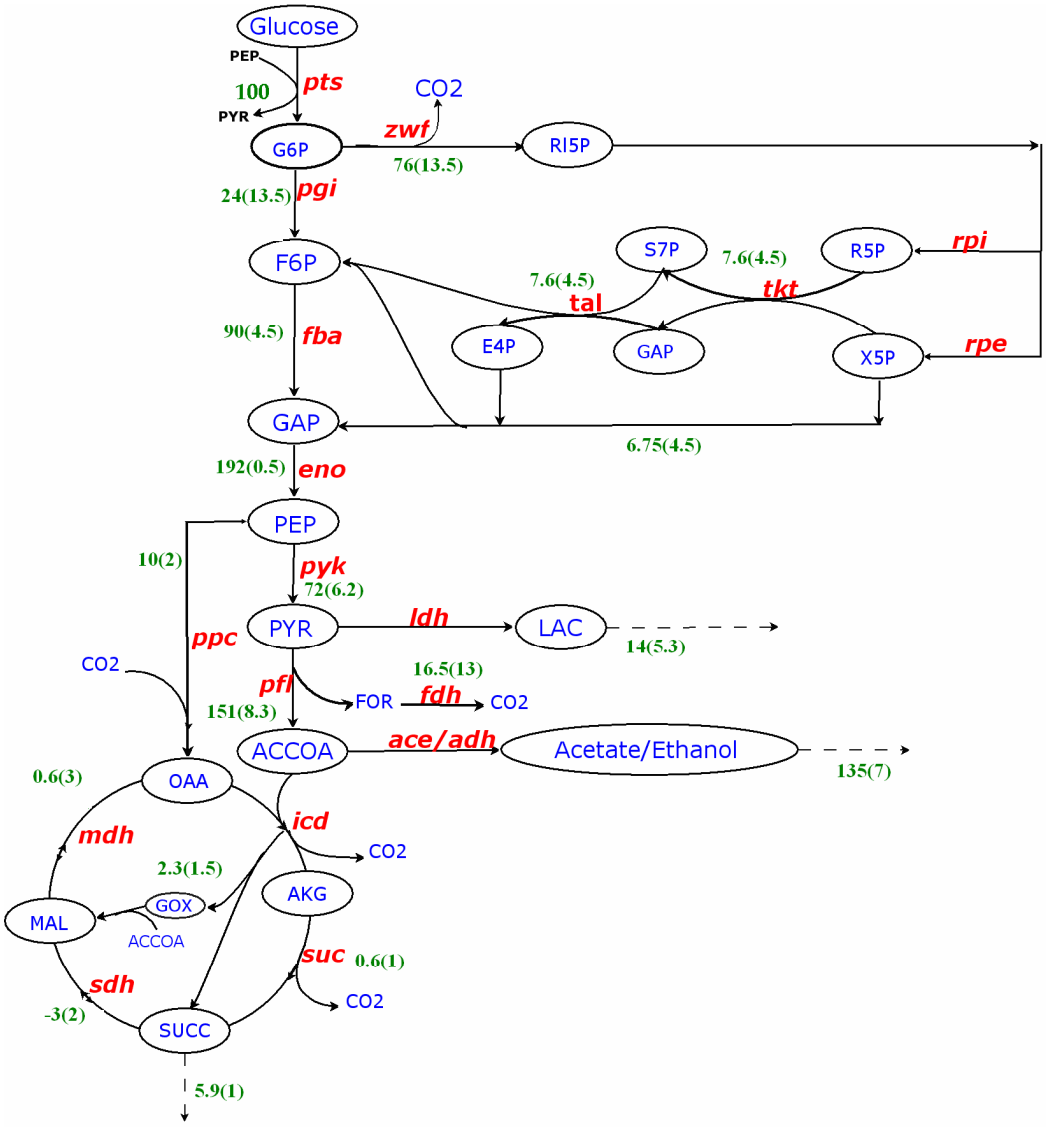

Figure 4.3: The metabolic flux map of W3110 *E coli* under anaerobic conditions obtained via ^{13}C flux analysis with 10% U-^{13}C glucose

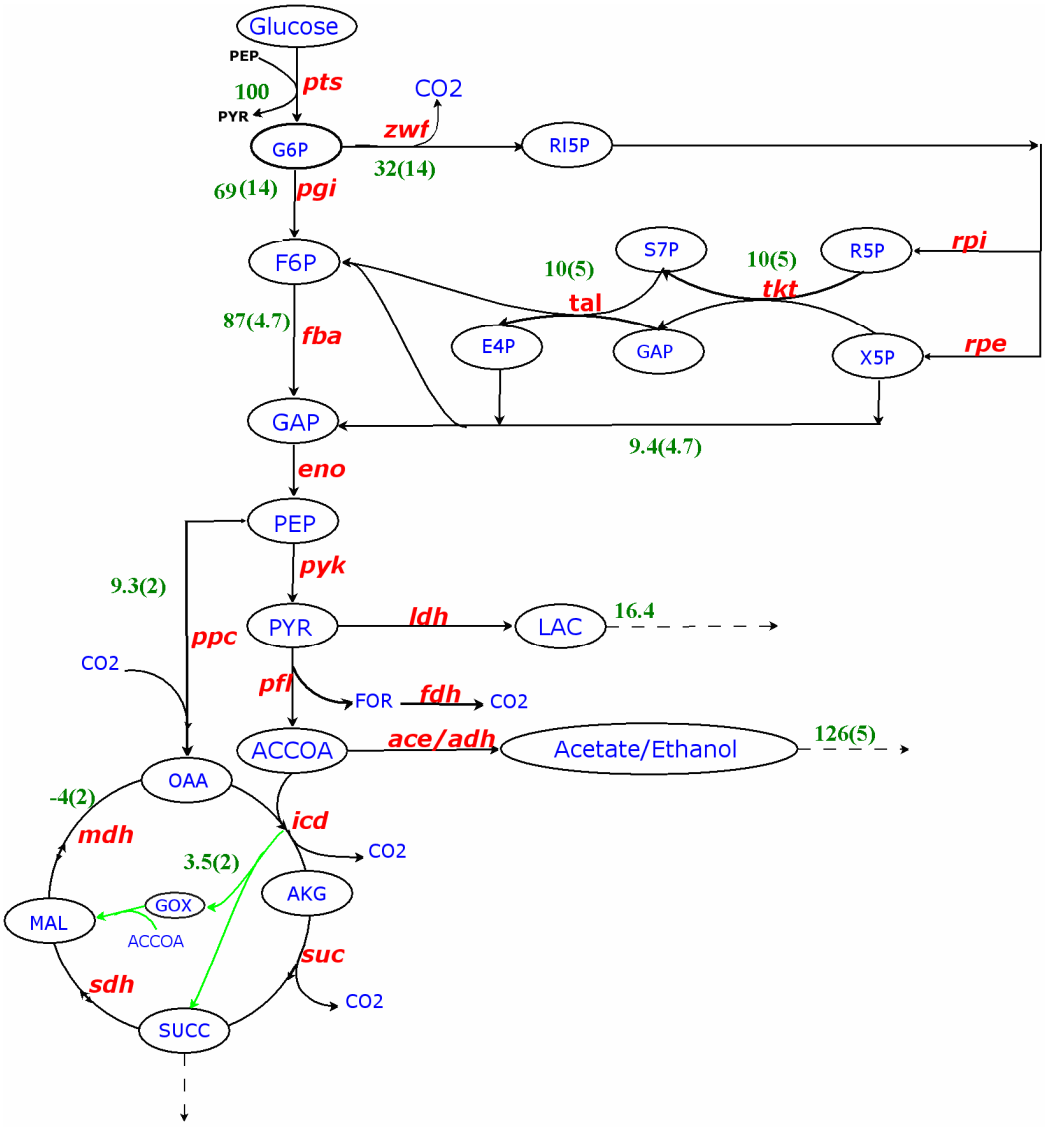

Figure 4.4 The metabolic flux map of W3110 *E coli* obtained via ^{13}C flux analysis with U-^{13}C glucose without acetate/ethanol measurement. Values in braces denote standard deviation associated with the flux.

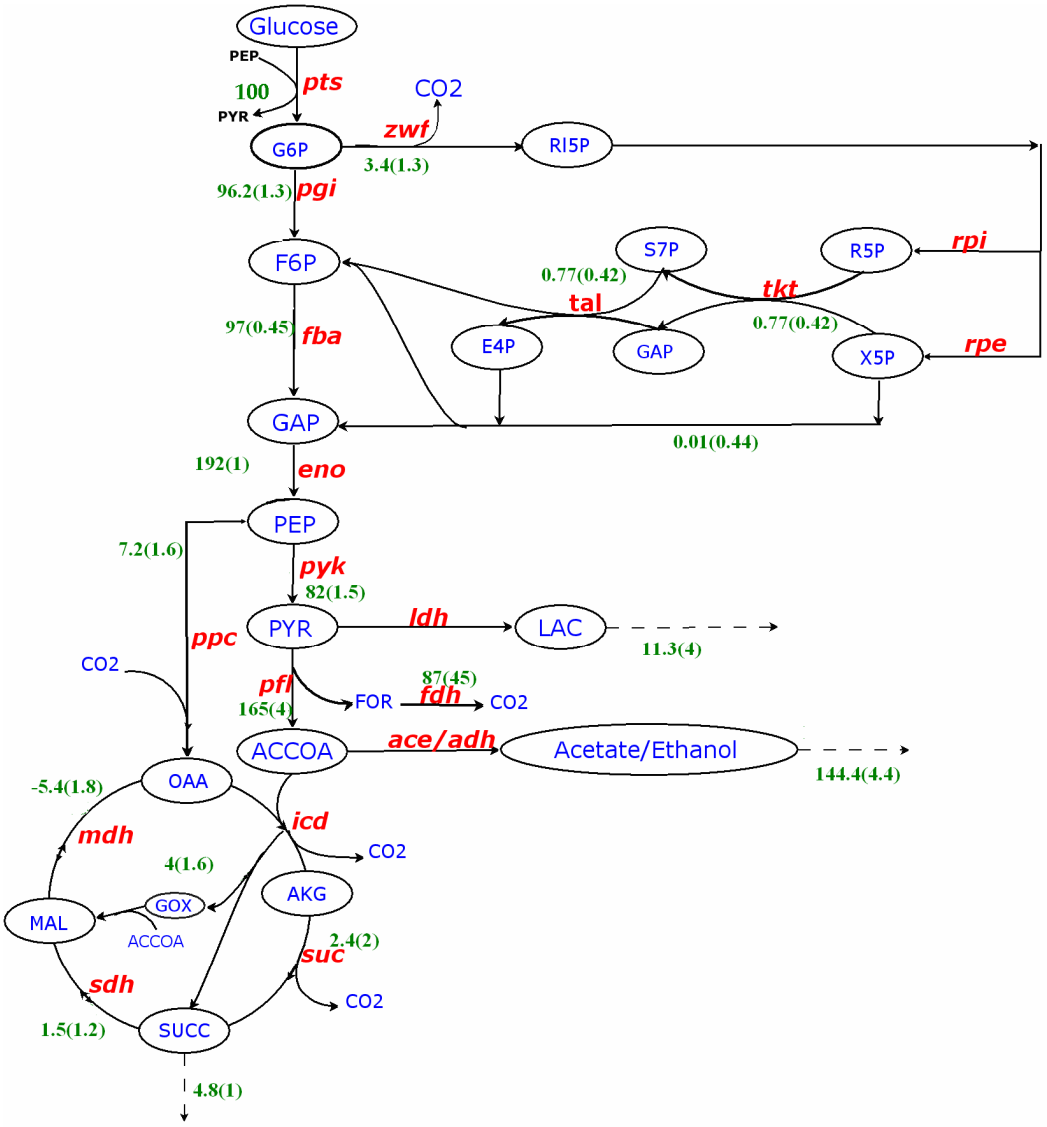

Figure 4.5: The metabolic flux map of W3110 *E coli* obtained via ^{13}C flux analysis with 10% U-^{13}C glucose and 25 % 1-^{13}C glucose without acetate/ethanol measurement. Values in braces denote standard deviation associated with the flux.

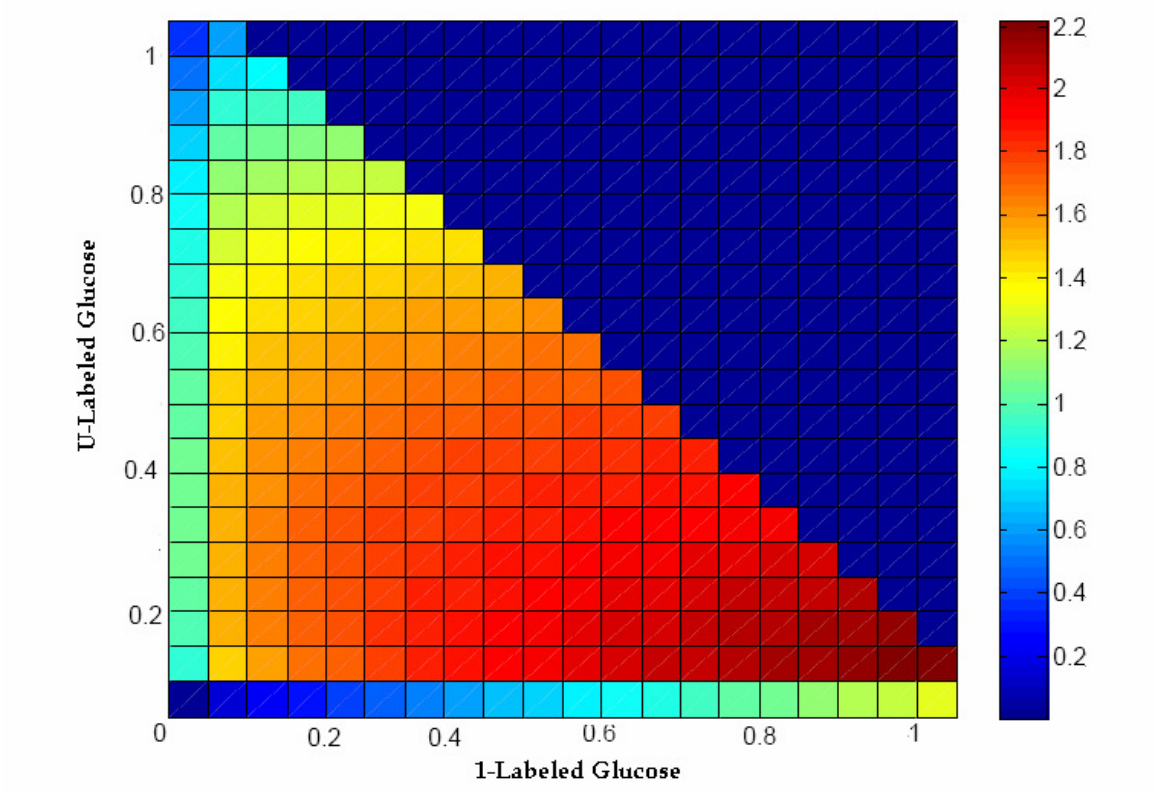

Figure 4.6: Optimal experimental design for metabolic flux analysis in *E. coli* under anaerobic condition grown on glucose. The information content (*IC*) is shown relative to the reference experiment [10% U-13C].

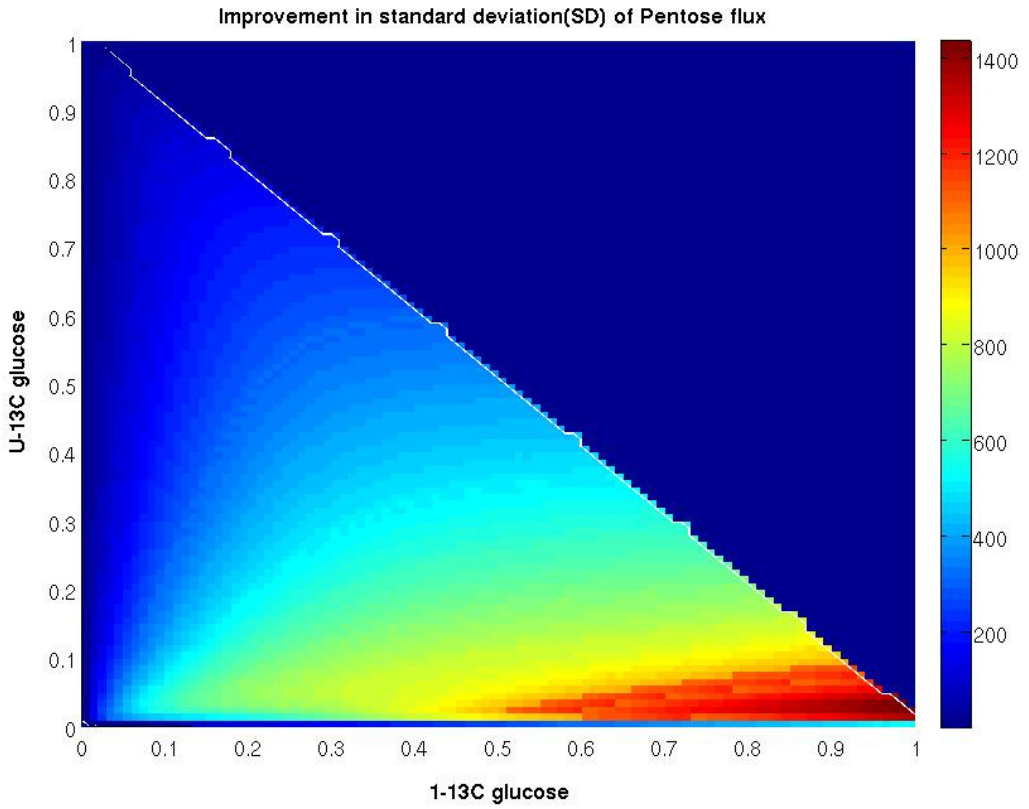

Figure 4.7: Optimal experimental design for metabolic flux analysis in *E. coli* under anaerobic condition grown on glucose. The information content (*IC*) is shown relative to the reference experiment [10% U-13C].

U-labeled glucose	1-labeled glucose	Average	SD	*IC*
0.10	0.00	23.00	14.00	1.00
0.10	0.10	4.77	0.34	1.30
0.10	0.20	4.77	0.32	1.60
0.10	0.25	4.78	0.33	1.65
0.10	0.30	4.78	0.28	1.70
0.10	0.35	4.78	0.32	1.72
0.10	0.40	4.76	0.27	1.73
0.10	0.45	4.80	0.28	1.73
0.10	0.50	4.78	0.32	1.74

Table 4.2:The flux and standard deviation estimated by NMR2Flux. All fluxes are reported relative to 100 moles of glucose consumed.

4.5 References

1. Bailey, J.E., *Toward a science of metabolic engineering.* Science, 1991. **252**(5013): p. 1668-1675.
2. Stephanopoulos, G., *Metabolic fluxes and metabolic engineering.* Metabolic Engineering, 1999. **1**: p. 1-11.
3. Schmidt, K., J. Nielsen, and J. Villadsen, *Quantitative analysis of metabolite fluxes in Escherichia coli, using two-dimensional NMR spectroscopy and complete isotopomer models.* Journal of Biotechnology, 1999. **71**: p. 175-190.
4. Fischer, E. and U. Sauer, *Metabolic flux profiling of Escherichia coli mutants in central carbon metabolism using GC-MS.* European Journal of Biochemistry, 2003. **270**(5): p. 880-891.
5. Fischer, E. and U. Sauer, *A novel metabolic cycle catalyzes glucose oxidation and anaplerosis in hungry Escherichia coli.* Journal of Biological Chemistry, 2003. **278**(47): p. 46446-46451.
6. Fischer, E., N. Zamboni, and U. Sauer, *High-throughput metabolic flux analysis based on gas chromatography-mass spectrometry derived 13C constraints.* Analytical Biochemistry, 2004. **325**(2): p. 308-316.
7. Emmerling, M., et al., *Metabolic flux responses to pyruvate kinase knockout in Escherichia coli.* Journal of Bacteriology, 2002. **184**(1): p. 152-164.
8. Szyperski, T., *Biosynthetically directed fractional ^{13}C-labeling of proteinogenic amino acids. An efficient analytical tool to investigate intermediary metabolism.* European Journal of Biochemistry, 1995. **232**: p. 433-448.

9. Ingraham, J.L., Maaloe, O. and Neidhardt, F.C.,, *Growth of the Bacterial Cell*. 1983, Sunderland, MA.: Sinauer Associates, .

10. Sauer, U., et al., *Metabolic Flux Ratio Analysis of Genetic and Environmental Modulations of Escherichia coli Central Carbon Metabolism*, in *The Journal of Bacteriology*

J. Bacteriol. 1999. p. 6679-6688.

11. Johnson, B.A. and R.A. Blevins, *NMRView: a computer program for the visualization and analysis of NMR data.* Journal of Biomolecular NMR, 1994. **4**: p. 603-614.

12. van Winden, W., et al., *Innovations in Generation and Analysis of 2D [13C,1H] COSY NMR Spectra for Metabolic Flux Analysis Purposes.* Metabolic Engineering, 2001. **3**(4): p. 322-343.

13. Sriram, G., et al., *Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional ^{13}C labeling, two-dimensional [^{13}C, ^{1}H] nuclear magnetic resonance, and comprehensive isotopomer balancing.* Plant Physiology, 2004. **136**: p. 3043-3057.

14. Shanks, G.S.J.V., *Flux identifiability-based optimal design of ^{13}C lableing experiment for glycine max (soybean) embryo metabolism.* Biotechnology Progress. **Manuscript in preparation**.

15. Ferenci, T., *Hungry bacteria: Definition and properties of a nutritional state.* Environmental Microbiology, 2001. **3**(10): p. 605-611.

16. Sauer, U., et al., *The Soluble and Membrane-bound Transhydrogenases UdhA and PntAB Have Divergent Functions in NADPH Metabolism of Escherichia coli.* J. Biol. Chem., 2004. **279**(8): p. 6613-6619.

17. Sauer, U. and B.J. Eikmanns, *The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria.* FEMS Microbiology Reviews, 2005. **29**(4): p. 765-794.

18. Dauner, M., J.E. Bailey, and U. Sauer, *Metabolic flux analysis with a comprehensice isotopomer model in Bacillus subtilis.* Biotechnology and Bioengineering, 2000. **76**: p. 144-156.

19. Schmidt, K., et al., *Quantification of intracellular metabolic fluxes from fractional enrichment ^{13}C-^{13}C coupling constraints on the isotopomer distribution in labeled biomass components.* Metabolic Engineering, 1999. **1**: p. 166-179.

5 Metabolic Flux analysis of *Escherichia coli ptsG* Mutant and Wild Type Consuming Glucose/Xylose under Anaerobic Conditions

5.1 Introduction

Plant biomass in the form of lignocellulosic material is the most abundant source of fermentable carbohydrates in the world [1]. It can be widely used to produce sustainable biobased products and fuels to replace depleting fossil fuels. The hydrolysis of lignocelluloses yields a mixture of pentose and hexose sugar which can be fermented to ethanol[1, 2]. Therefore, a biomass to ethanol process requires an organism which ferments the multiple sugars. *Escherichia coli* unlike yeast which is commonly used in ethanol fermentation, has natural ability to consume both pentose and hexose both under aerobic and anaerobic conditions. It has also been engineered to selectively produce ethanol at high yield from mixed sugars [3].

However, due to carbon catabolite repression (CCR), there is a sequential consumption of sugars. Glucose is consumed first and the consumption of other pentose sugars is delayed and often incomplete resulting in lower yield. Hence, the development of *E. coli* strain capable of consuming both glucose and pentose sugars like xylose at the same time, would beneficial for the making the process economically feasible. Such a strain is also valuable for the production of other bulk chemicals like lactic acids, acetic acids and succinic acids[4].

PTS is the main glucose transport system and it consists of three cytosolic proteins E-I, HPr, IIA and IIBC [5]. The glucose transport through PTS leads to

dephosphorylation of IIA protein. Phosphorylated IIA protein activates adenylate cyclase (AC) which converts ATP to cAMP. The glucose exerts CCR by depriving the cell of CRP-cAMP required for the synthesis of proteins in the catabolism of pentose sugars [5]. In the PTS mutant, where glucose is not transported by PTS, is likely not to show CCR. Different groups have explored strategies to disrupt CCR by inactivating PTS components. Ingram *et al* [6] reported the use of fosfomycin to select a PTS mutant from the ethanologenic strain of *E. coli*. Although they reported higher ethanol yield on xylose, CCR was still observed resulting in incomplete utilization of xylose. A *ptsG* mutant has been found to simultaneously consume glucose and xylose[4]. Flores *et al* [7, 8] also characterized a PTS-Glc mutant isolated from glucose-limited continuous culture of a PTS- mutant. Although, concurrent consumption of glucose and arabinose was observed, the consumption of xylose was delayed.

There have been few efforts to characterize *E. coli* co-utilizing glucose and xylose, but these studies are based on study the effect to "lumped response" such as growth rate, substrate uptake rate and product yield [7,8,6]. The prediction of how cells will respond to a genetic modification is difficult due to the complexity of metabolic kinetics and regulation operating in a large number of reactions. Hence, in order to analyze the extent to which cell physiology has been changed and to gain insights for future modification, analysis should involve the details of pathways of interest as well as of interacting pathways. Since, CCR is global phenomena involving global regulators like CRP-cAMP and Mlc, any perturbation in PTS leads to pleiotropic effects. Hence, it requires a systems level approach like MFA and transcriptomics.

Systems level analysis has been extensively used to characterize yeast, which is another promising candidate for conversion of biomass into ethanol. Unlike *E. coli*, wild type yeast is not able consume xylose, the most important pentose sugar in lignocellulosic hydrolyzates. Xylose genes encoding xylose reductase and xylitol dehydrogenase from *Pichia stipitis* were expressed and using a systems level approach like MFA and transcriptomics analysis, redox balance was found to be the main bottleneck in consumption of xylose both glucose and xylose [9].

In *E. coli*, the most of the engineering efforts to understand the pts mutant was directed to increase the metabolic availability of PEP for synthesis of aromatic compounds. Chen *et al* [10] characterized the effect of PTS inactivation through NMR studies and predicted system wide effects on cell physiology. Still, there has been no report of usage of system wide tool like DNA microarray or ^{13}C MFA to characterize pts mutants.

With the development of ethanologenic *E. coli* and its natural ability ferment glucose and xylose under anaerobic conditions, *E. coli* is a promising candidate for metabolic engineering for the production of ethanol from lignocellulosic biomass. In this work, we are going to present the ^{13}CMFA analysis of wild type and *ΔptsG* mutant of *E. coli* under anaerobic conditions grown on a mixture of glucose and xylose.

5.2 Material and Methods

Bacterial strains and cultural conditions

Wild-type strain of W3110 (ATCC#27325) *Escherichia coli* K12 strain were obtained from the American Type Culture Collection (Manassas, VA). Deletion mutants *ΔptsG*, was prepared from wild type *E. coli* using a one step gene inactivation method

according to Datsenko *et al*[11]. The strains from glycerol stock at -80C. were steaked onto LB plates and were incubated for 12 hours at 37C. The colonies from the LB plates were used to inoculate 3 hunge tubes filled with LB media containing 5g/L of xylose. The tubes were incubated until OD reached 0.4. This actively growing cells were used to inoculate 24 hunge tube filled with minimal media supplemented with 5g/LThe wild type and *ΔptsG* mutant were grown in MOPS (4-Morpholinepropanesulfonic acid) minimal medium with 1.32mM Na_2HPO_4 in place of K_2HPO_4 with glucose and xylose as the carbon source. The minimal media was also supplemented with 1g/l of $NaHCO_3$. The initial glucose and xylose concentration was 5g/L. The ^{13}C labeling experiment was carried out at 37^{o}C in SixFors multi-fermentation system (Infors HT, Bottmingen, Switzerland) with six 400 mL working volume fermenters and independent control of temperature, pH, and stirrer speed (200 rpm) with three replicates each for w3110 and *ptsG* mutant. The fermentation system is computer controlled using manufacturer IRIS NT software. Each fermenter is fitted with a condenser to prevent evaporation, which was operated with a 0˚C cooling methanol–water supply. Anaerobic conditions were maintained by flushing the headspace with ultrahigh purity argon (Matheson Tri-Gas, Inc., Houston, TX) at 0.01 LPM. An oxygen trap (Alltech Associates, Inc., Deerfield, IL) was used to eliminate traces of oxygen from the gas stream. To maintain sterile conditions, 0.2 mm and 0.45 mm HEPA filters (Millipore, Billerica, CA) were used to fit the inlet and outlet lines, respectively.

The wild type labeling experiment was carried out with 10% U-^{13}C glucose, 25% 1-^{13}C glucose and 65% naturally labeled glucose and unlabeled xylose. For *ΔptsG* mutant, 50%U-^{13}C glucose, 10% 1-^{13}C glucose and 40% naturally labeled glucose with

10% U-^{13}C xylose, 10% 1-^{13}C xylose and 80% naturally labeled xylose was used as carbon substrate.

Analytical procedures

Dry cell weight was monitored by measuring optical density (OD_{550}) using spectrophotometer (Genesys 20, Thermospectronic, and Madison, Wisconsin) (1 OD=0.36 g DW/L). Glucose and fermentation products were measured on Waters (Milford, MA) HPLC system using Aminex column(HPX-87H, Bio-Rad, Hercules, CA, USA).

Preparation of cellular amino acids for 2D NMR measurement

For isotopomer analysis by NMR, 350 mL of biomass was harvested at OD_{550} of 0.4 and the cell growth was quenched by keeping the cells on ice water. The cells were first centrifuged at 5000g for 15 min at +4˚C. The pellet were washed with 0.9% saline water and centrifuged again for 15min at 5000 g. About 30mg of biomass(estimated dry weight) was transferred in four hydrolysis tubes (Pierce Endogen, Rockford, IL), to which 6 N hydrochloric acid(Pierce Endogen, Rockford, IL) was added in the 1 mL of HCl:4 mg of biomass. The hydrolysis tube was evacuated, flushed with nitrogen to remove residual oxygen, and reevacuated. The hydrolysis was performed at 110°C for 12 hrs. The acid in the hydrolyzates was evaporated in a Rapidvap evaporator (Labconco, Kansas City, MO). The residue was reconstituted in 2 ml of deionized water, lyophilized for 72 h, and finally the dissolved in 500 μL of D_2O(Sigma, St. Louis) in an NMR tube. The pH of each NMR sample was adjusted to 1 using DCl (Sigma, St. Louis). Separate NMR sample was

prepared for each replicates and NMR spectroscopy experiments were carried out for each of them.

NMR experiment

Two-dimensional [^{13}C, ^{1}H] HSQC NMR spectra were collected on a Bruker Avance DRX 700 MHz spectrometer (Bruker Instruments, Billerica, MA) at 298 K. The reference to 0 ppm was set using the methyl signal of dimethylsilapentanesulfonate (Sigma, St. Louis) as an internal standard. The resonance frequency of ^{13}C and ^{1}H were 174.7 MHz and 700 MHz, respectively. The spectral width was 5,482.26 Hz along the ^{1}H (F2) dimension and 3916.13 Hz along the ^{13}C (F1) dimension. Peak aliasing was used in order to minimize the sweep width along the F1 dimension. The number of complex data points was 1,024 (^{1}H) x 900 (^{13}C). A modification of the INEPT (insensitive nuclei enhanced by polarization transfer) pulse sequence was used for acquiring HSQC spectra. The number of scans was generally set to 8.

The software Xwinnmr (Bruker Instruments, Billerica, MA) was used to acquire all spectra, and the software NMRView[12] was used to quantify nonoverlapping multiplets on the HSQC spectrum. Overlapping multiplets (α-amino acid), which could not be processed with NMRView, were quantified by a peak deconvolution software based on spectral processing algorithm proposed by Van Winden *et al* [13].

Metabolic flux calculation

Fluxes were evaluated from isotopomers data by using generic software NMR2Flux developed previously in our group by Sriram *et al* [14]. NMR2Flux uses isotopomer balancing and a global optimization routine to minimize the difference

between simulated and experimental intensities. The objective of this flux evaluation procedure is to evaluate a set of stoichiometrically feasible fluxes that best accounts for the measured isotopomer abundances and extracellular flux measurements. Multiple simulations were carried out from different starting point in order to verify the global minimum. The statistical error analysis was performed by Monte Carlo simulation. Flux analysis was carried out for each replicates separately. The fluxes were estimated as the average of flux values of the replicates.

5.3 Results and Discussion

Physiological analysis

In order to determine the effect of *ptsG* mutation on the utilization of sugars, batch anaerobic fermentation of wild type and *ptsG* mutant was carried out in 500mL fermenters with glucose and xylose as the carbon source. The *ptsG* mutant didn't grow in the absence of bicarbonate in the medium. Hence, bicarbonate which provide CO_2, was added in both wild type and *ΔptsG* culture. CO_2 has been found to reduce the lag in the fermentation [15]. It is needed for the synthesis of oxaloacetate (OAA) from phosphoenolpyruvate (PEP) by anaplerotic reaction catalyzed by phosphoenolpyruvate carboxylase (ppc). In initial culture, the amount of CO_2 needed is not enough to support maximal growth. This is more valid in anaerobic culture as the partial oxidation of sugars results in less CO_2 production. Hence, the addition of CO_2 to the medium reduces the lag phase by providing CO_2 needed for the biosynthesis. The *ptsG* mutant doesn't produce enough CO_2 to support growth. Hence, bicarbonate is essential to supply CO_2.

The experimentally determined growth parameters are summarized in Table 5.1. The maximum specific growth rate and growth yield for wild type was found to be

$0.3hr^{-1}$ and 0.09±0.0035 g biomass/g of sugar and is consistent with previous reported for JM101 wild type strains of *E. coli* under anaerobic conditions[16]. On the other hand, *ptsG* mutant grew more slowly and exhibited lower biomass yield. *ptsG* mutant metabolized glucose and xylose simultaneously while the wild type consumed xylose only after glucose has been completely consumed. The cometabolism of sugars in *ptsG* mutant can be explained by the fact that in the absence of active PTS system, the glucose is transported via by galactose transporter (galP) and subsequently phosphorylated to glucose-6-phosphate (G6P) by glucokinase (glk). Hence, the IIA protein exists mainly in phosphorylated form and hence it can activate adenylate cyclase(AC) which converts ATP to cAMP level. The increased level of cAMP in Δ*ptsG* results in the expression xylose of genes even in the presence of glucose.

In wild type, there was a lag phase of 14 hours after the depletion of glucose before xylose consumption started. Interestingly, the lag phase lasted until all formate in the medium is exhausted; the physiological significance of this finding is not known at present. The formate is known to be toxic to the cells as it can penetrate the cell and cause cell damage [17]. However, its conversion to CO_2 by formate hydrogenlyase (fhl) seems unlikely because it is activated at low pH. It must be degraded by some other enzymes.

The complete fermentation with the wild type took 40 hrs and glucose and xylose were consumed completely. The *ptsG* mutant fermentation required 60 hours and both glucose and xylose were completely utilized. The initial glucose uptake in wild type was found to be 4.3 mmol/ gm hr while ptsG inactivation resulted in lower glucose uptake rate of 0.63. The cell still can transport glucose through galP or the mannose PTS, but

with lower efficiency and slower rate. The total initial carbon uptake was found to be 25 mmol / gm hr and 19.5 mmol/gm hr in wild type and *ptsG* mutant respectively.

Optimal labeling substrate mixture

The choice of substrate is very important design criteria in ^{13}C flux analysis as it affect the flux identifiability i.e. whether a flux can be estimated from the labeling mixture or not. Previously conducted experiment with 10% U-^{13}C glucose as the sole labeling carbon has enabled us in accurate flux analysis in wild type *E. coli* but flux through pentose pathway was not well resolved. Using identifiability analysis, optimal mixture of 10% U-^{13}C glucose and 25% 1-^{13}C glucose was found which enabled us to correctly determine pentose fluxes. This prompted us to use same identifiability analysis to select optimal mixture for both wild type and *ptsG* mutant.

The uptake of CO_2 from bicarbonate can result in dilution of intracellular CO_2 pool produced by 6-phosphogluconate dehydrogenase (gnd), pyruvate dehydrogenase formate hydrogenlyase and consumed by anaplerotic reaction by ppc to produce OAA which is a precursor for many amino acids. Hence, it is crucial to take into the account of impact of CO_2 on flux identifiability. The flux estimate from the previous experiment was used to explore various labeling mixture. The 10% U-^{13}C and 25%1-^{13}C glucose was chosen as optimal substrate for the wild type labeling experiment.

On the other hand, the situation for the *ptsG* mutant is more complicated. In addition to uptake of carbon dioxide from bicarbonate, it also utilizes both xylose and glucose simultaneously. The xylose uptake is approximately five times the glucose uptake rate which makes flux estimation problematic. Additionally, the ^{13}C labeled xylose is 10 times more expensive than U-^{13}C glucose. The amount of labeling

incorporated is also important as it affect the noise associated with NMR measurements. Hence, the labeling mixture is crucial for flux estimation. Identifiability analysis was carried with the assumed set of fluxes in *ptsG* mutant to explore different labeling substrate as shown in Table 5.2. The labeling mixture of 50% U-^{13}C glucose, 10% each of 1-^{13}C glucose, 1-^{13}C xylose and U-^{13}C xylose was found to be optimal. This is consistent with previous finding by Peterson *et al* [18], who reported that it is important to use labeled cosubstrate to accurately estimate of metabolic fluxes in *C. glutamicium* coutilizing lactate and glucose.

Metabolic flux analysis

In order to gain insight into effect of *ptsG* mutation on the intracellular metabolism of the sugars, metabolic flux analysis was carried out in wild type and *ptsG* mutant. The previously constructed metabolic model was extended to incorporate xylose uptake by high affinity system xylFGH and subsequent phosphorylation to xylose 5 phosphates. The glucose uptake in *ptsG* was modeled as initial uptake by galP and subsequent phosphorylation by glk to glucose 6 phosphates (G6P).

The phosphoenolpyruvate carboxylase (ppc) consumes CO_2 as bicarbonate, but bicarbonate cannot diffuse through the cell wall. Hence, it decomposes to CO_2 which is then transported inside the cell and then again converted back to bicarbonate to be used up by the cell. Hence, the CO_2 uptake form bicarbonate is modeled as bicarbonate $\rightarrow CO_2$ to the intracellular pool of CO_2. The ED pathway which converts glucose-6-phosphate (G6P) to pyruvate (PYR) and glycerol 3-phosphate (G3P) was added to the model as it may be active in the absence of CCR.

The NMR2Flux was used to estimate metabolic fluxes from the labeling data and extracellular measurements. Except for his-β, Pr-δ, Arg-δ and His-α, there was good match between simulated and experimental intensities. Pr-δ, Arg-δ reflects labeling pattern of same precursor and thus one would similar intensities for the them. The fact their intensities were comparable suggests that the misfit is not due to the error in NMR intensities but due to the metabolic network model used. Hence, these were not considered into the flux estimation. His-β was found to have long range coupling, which concurs with previous finding by Szyperski *et al* [19]. The spectral deconvolution of the peak didn't result in the reliable estimate of multiplet intensities. Hence, it was not included in flux estimation

Our flux analysis framework allows us to test various networks. Malic enzyme (mez) reactions which supply cells with NADH and converts MAL to PYR was added to the model. The addition of mez into the metabolic network didn't change the significantly affect the chi-square (900 to 850), but the standard deviation of fluxes in TCA cycle increased. Hence, it was not included in the model.

Figure 5.4 shows the intracellular flux distribution in wild type and its *ptsG* mutant. Fluxes are reported relative to 600 moles of carbon taken (100 moles of glucose in wild type.). The relative fluxes are also reported in appendix B1.

The glycolytic flux through enolase (eno) catalyzing conversion of G3P to PEP was approximately identical both wild type (194±1.1) and *ptsG* mutant (193±3). Entner-Doudoroff (ED) pathway which is an another route for conversion of G6P to pyruvate (PYR) and glycerol 3-phosphate(G3P) bypass the enolase, was found to be inactive in the wild type(0.25±1) while the flux through ED in *ptsG* mutant was found to be 0.64±1.7.

Hence, we cannot infer about its activity in the *ptsG* mutant. The flux through glyoxalate shunt which converts isocitrate (ICT) to malate was found to be 11±4.5 and 5±4.5. Hence, it is most likely inactive both wild type and *ptsG*.

Pentose pathway: The pentose pathway provides cells with NADPH and precursors (Ribose 5-phosphate and Erythrose 4-phosphate) needed for biosynthesis. Under anaerobic conditions, it may also serve as source for CO_2 needed for biosynthesis. The flux through oxidative part of pentose pathway (zwf) was found to be 2.5±1.7 while in *ptsG* mutant it was 11.7±6.6. Similar to our results, Flores *et al* [8] found the relative flux through oxidative part of pentose pathway (zwf) increases substantially as result of PTS inactivation under aerobic case. The flux through non-oxidative part of pentose pathway (transketolase and transaldolase) was higher in *ptsG* mutant in accordance with xylose metabolism metabolism via pentose pathway. Gene expression of pentose enzymes does not change in *ptsG* grown on glucose-xylose mixture [20]. Thus, the pentose gene is not regulated by CRP-cAMP. Furthermore, it can be concluded the low flux in the wild type is due to low availability of pentose rather than gene expression.

The *ptsG* mutant exhibited much higher flux through pyruvate kinase (pyk). Since in the wild type, glucose transport through the PEP dependent PTS is coupled with conversion of PEP to PYR. While in *ptsG* mutant, both xylose and glucose are transported via PEP independent system, hence all PEP metabolisms to pyruvate are through pyruvate kinase (pyk).

ATP balance

The flux distribution was used to the amount of ATP production. We found net ATP production in central carbon metabolism was 235 whereas it was 133 in the *ptsG*

mutant. The lower ATP in the *ptsG* mutant is mainly due to the inefficient transport of sugars in *ptsG* mutant. In wild type, glucose is transported via PTS and it enters cells as glucose 6 phosphate consuming one phosphate bond in the form of PEP. On the other hand in the *ptsG* mutant, both glucose and xylose is transported via ATP dependent system and needs to be phosphorylated before its utilization by EMP pathway and pentose pathway respectively. Hence, transport and activation of each molecule of sugars consumed 2 moles of ATP compared to 1 mole in the case wild type.

The unregulated uptake of non-PTS sugar by an externally added phosphorylated sugar or cAMP was shown to cause cell death as a result of accumulation of methylglyoxal (MG), a toxic metabolite produced from dihydroxyacetone phosphate(DHAP)[]. The MG accumulation does not occur on growth on glucose as aldose catalyzing the reaction fructose1-6→ glycerdehyde-3-phosphate (G3P) + DHAP has low activity. On other hand, growth on xylose bypass aldose through PP pathway. Since, the flux through pentose is much higher in *ptsG* mutant, MG synthesis may occur via triose phosphate isomerae(G3P→ DHAP) in the *ptsG* mutant. The flux through methylglyoxal bypass which converts DHAP to pyruvate, could not be estimated by our approach as it is result in the same carbon rearrangement as glycolysis. Hence, we cannot rule out slow growth of *ptsG* mutant due to methylglyoxal synthesis in the mutant.

Since our analysis revealed the slow growth of *ptsG* mutant is due to less efficient transport of xylose compared to glucose, so one possible way to increase the growth of the *ptsG* mutant is to engineer efficient transport of xylose which will consume just one mole of ATP per mole xylose-5-phosphate formed from xylose. Another possible approach is to engineer efficient transport of glucose which is as efficient as PTS.

5.4 Conclusion

The effect of *ptsG* mutation was studied by carrying out 13C labeling experiment on sugar mixture of wild type and *ptsG* mutant. It was found the catabolite repression was found in the wild type while in *ptsG* mutant, catabolite repression was not observed resulting simultaneous consumption of glucose and xylose. Further analysis revealed that in wild consumed glucose much faster than the *ptsG* mutant. From the flux analysis, it was found that the slow growth of the *ptsG* mutant is primarily due to less amount of ATP available per 600 moles of glucose consumed. Less efficient transport of xylose and glucose results in lesser availability of ATP in *ptsG* mutant.

Since, methylglyoxal pathway results in same carbon rearrangement as glycolysis, the flux through MG pathway cannot be estimated. The flux through MG pathway is likely to be higher in the *ptsG* mutant as flux through PP pathway is much higher in the *ptsG* mutant.

5.5 **Reference**:

1. Himmel, M.E., M.F. Ruth, and C.E. Wyman, *Cellulase for commodity products from cellulosic biomass.* Current Opinion in Biotechnology, 1999. **10**(4): p. 358-364.
2. Percival Zhang, Y.H., M.E. Himmel, and J.R. Mielenz, *Outlook for cellulase improvement: screening and selection strategies.* Biotechnology Advances, 2006. **24**(5): p. 452-481.
3. Ihssen, J. and T. Egli, *Global physiological analysis of carbon- and energy-limited growing Escherichia coli confirms a high degree of catabolic flexibility*

and preparedness for mixed substrate utilization. Environmental Microbiology, 2005. **7**(10): p. 1568-1581.

4. Gosset, G., *Improvement of Escherichia coli production strains by modification of the phosphoenolpyruvate:Sugar phosphotransferase system.* Microbial Cell Factories, 2005. **4**: 14
5. Saier, M.H., Jr ;Ramseir,T.M; Reizer,J., *Regulation of carbon utilization*, in *Escherichia coli and Salmonella:Cellular and Molecular Biology*, F.C. Niedhart, Schaechter,M.; Curtiss III,R; Lin,E.C.C; Low,K.B; Magasanik,B.; Reznikoff,W.S.;, Editor. 1996, American Society for Microbiology: Washington,D.C. p. 1325-1343.
6. Lindsay, S.E., R.J. Bothast, and L.O. Ingram, *Improved strains of recombinant Escherichia coli for ethanol production from sugar mixtures.* Applied Microbiology and Biotechnology, 1995. **43**(1): p. 70-5.
7. Flores, N., et al., *Pathway engineering for the production of aromatic compounds in Escherichia coli.* Nature Biotechnology, 1996. **14**(5): p. 620-3.
8. Flores, S., et al., *Analysis of carbon metabolism in Escherichia coli strains with an inactive phosphotransferase system by 13C labeling and NMR Spectroscopy.* Metabolic Engineering, 2002. **4**(2): p. 124-137.
9. Pitkanen, J.-P., et al., *Metabolic flux analysis of xylose metabolism in recombinant Saccharomyces cerevisiae using continuous culture.* Metabolic Engineering, 2003. **5**(1): p. 16-31.
10. Chen, R., et al., *Metabolic consequences of phosphotransferase (PTS) mutation in a phenylalanine-producing recombinant Escherichia coli.* Biotechnol Prog 1997. **13**(6): p. 768-75.
11. Datsenko, K.A. and B.L. Wanner, *One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(12): p. 6640-6645.
12. Johnson, B.A. and R.A. Blevins, *NMRView: a computer program for the visualization and analysis of NMR data.* Journal of Biomolecular NMR, 1994. **4**: p. 603-614.
13. van Winden, W., et al., *Innovations in Generation and Analysis of 2D [13C,1H] COSY NMR Spectra for Metabolic Flux Analysis Purposes.* Metabolic Engineering, 2001. **3**(4): p. 322-343.
14. Sriram, G., *Flux analysis in central carbon metabolism in plants: ^{13}C NMR experiments and analysis*, in *Chemical Engineering*. 2004, Iowa State University: Ames. p. 231.
15. Repaske, R. and M.A. Clayton, *Control of Escherichia coli growth by CO2.* J Bacteriol FIELD Full Journal Title:Journal of bacteriology, 1978. **135**(3): p. 1162-4.
16. Sauer, U., et al., *Metabolic Flux Ratio Analysis of Genetic and Environmental Modulations of Escherichia coli Central Carbon Metabolism*, *J. Bacteriol.* 1999. p. 6679-6688.
17. Kirkpatrick, C., et al., *Acetate and Formate Stress: Opposite Responses in the Proteome of Escherichia coli.* J. Bacteriol., 2001. **183**(21): p. 6466-6477.

18. Petersen, S., et al., *In Vivo Quantification of Parallel and Bidirectional Fluxes in the Anaplerosis of Corynebacterium glutamicum.* J. Biol. Chem., 2000. **275**(46): p. 35932-35941.
19. Fiaux, J., et al., *13C NMR flux ratio analysis of Escherichia coli central carbon metabolism in microaerobic bioprocesses.* Journal of the American Chemical Society, 1999. **121**(6): p. 1407-1408.
20. Dharmadi, Y., *Metabolic Engineering of Escherichia coli for the Utilization of Plant Sugar Mixture*. 2005, Iowa State University: Ames.

Table 5.1: Anaerobic growth parameters of exponentially growing *E. coli* strains

	w3110	ptsG
Biomass Yield (g g-1)	0.1	0.05
Glucose uptake rate(mmol g-1 h-1)	4.3	0.63
Xylose uptake rate(mmol g-1 h-1)	0.98	3.15
Ethanol Yield(mol / mol)	0.72	0.8
Acetate Yield(mol / mol)	0.68	0.75
LactateYield(mol /mol)	0	0
Succinate Yield(mol / mol)	14.8	14.8
Formate Yield(mol / mol)	126	75

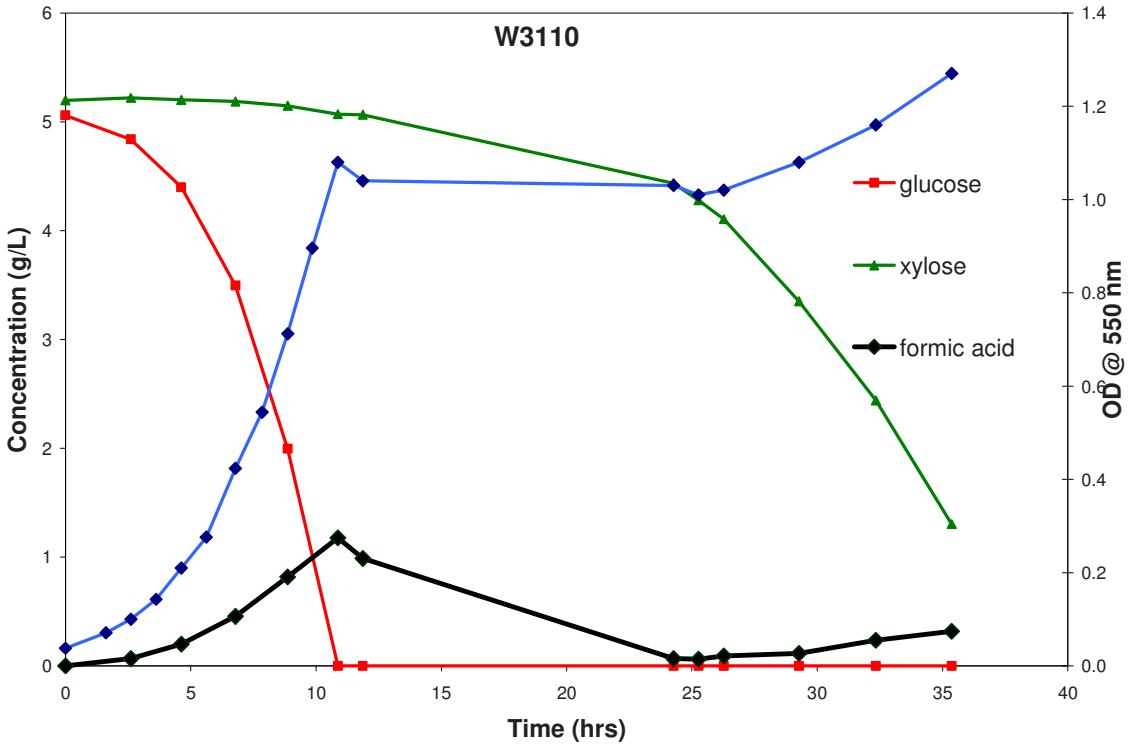

Figure 5. 1: Growth and sugar consumption profile of wild type (w3110). *E. coli* fermentation on minimum media containing 5g/L of glucose and 5g/L xylose**.**

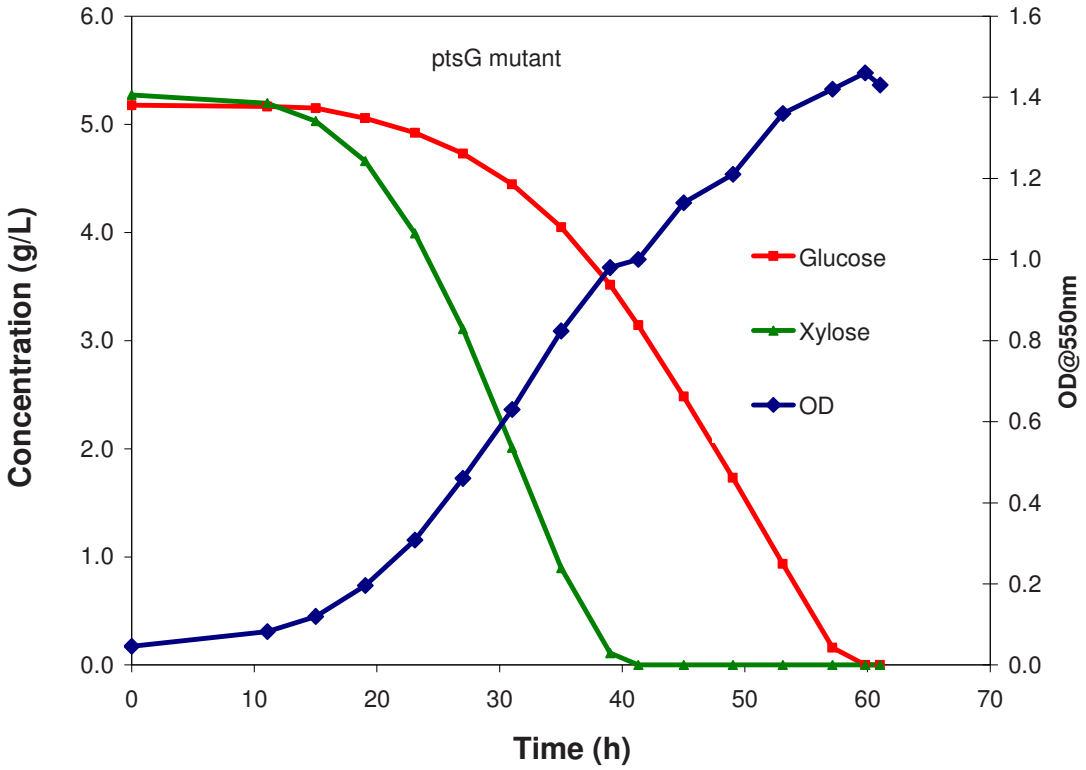

Figure 5.2: Growth and sugar consumption profile of *ptsG* mutant *E. coli* fermentation on minimum media containing 5g/L of glucose and 5g/L xylose**.**

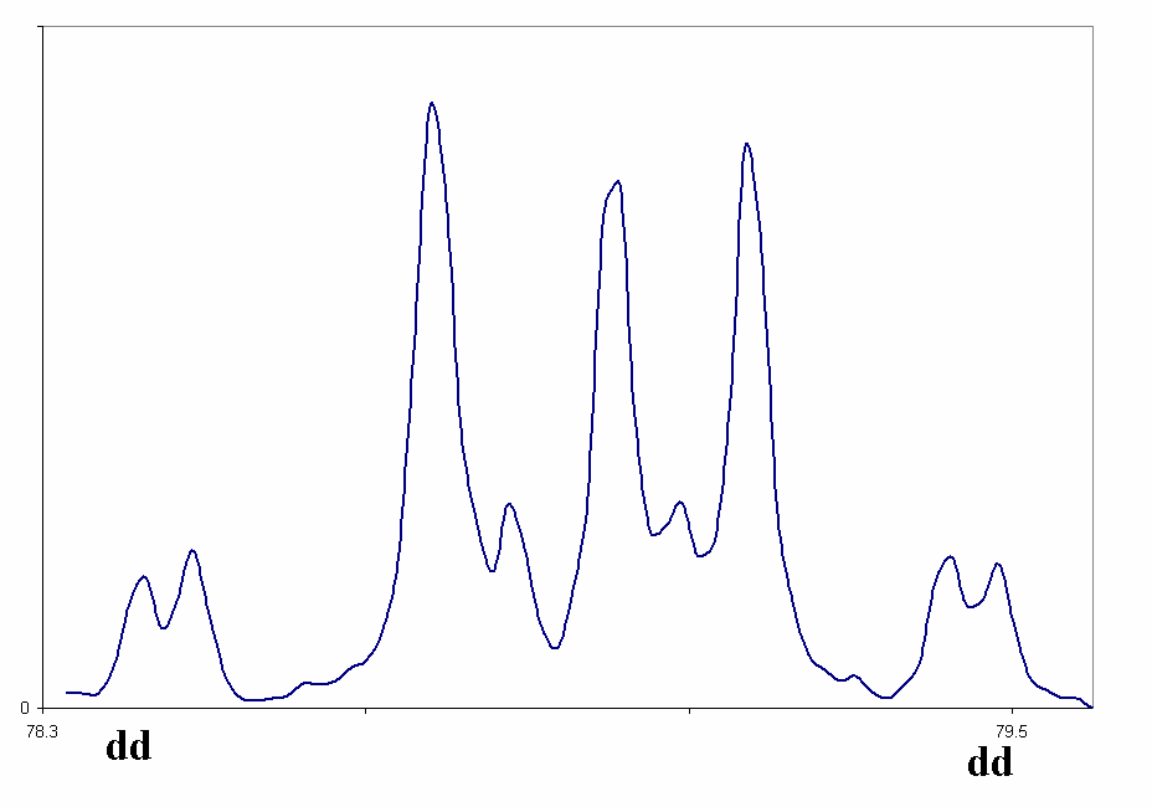

Figure 5.3: His-β peak on from *ptsG* protein hydrolyzates. Doublet of doublets arising from long range coupling between β and δ carbon of histidine can be seen

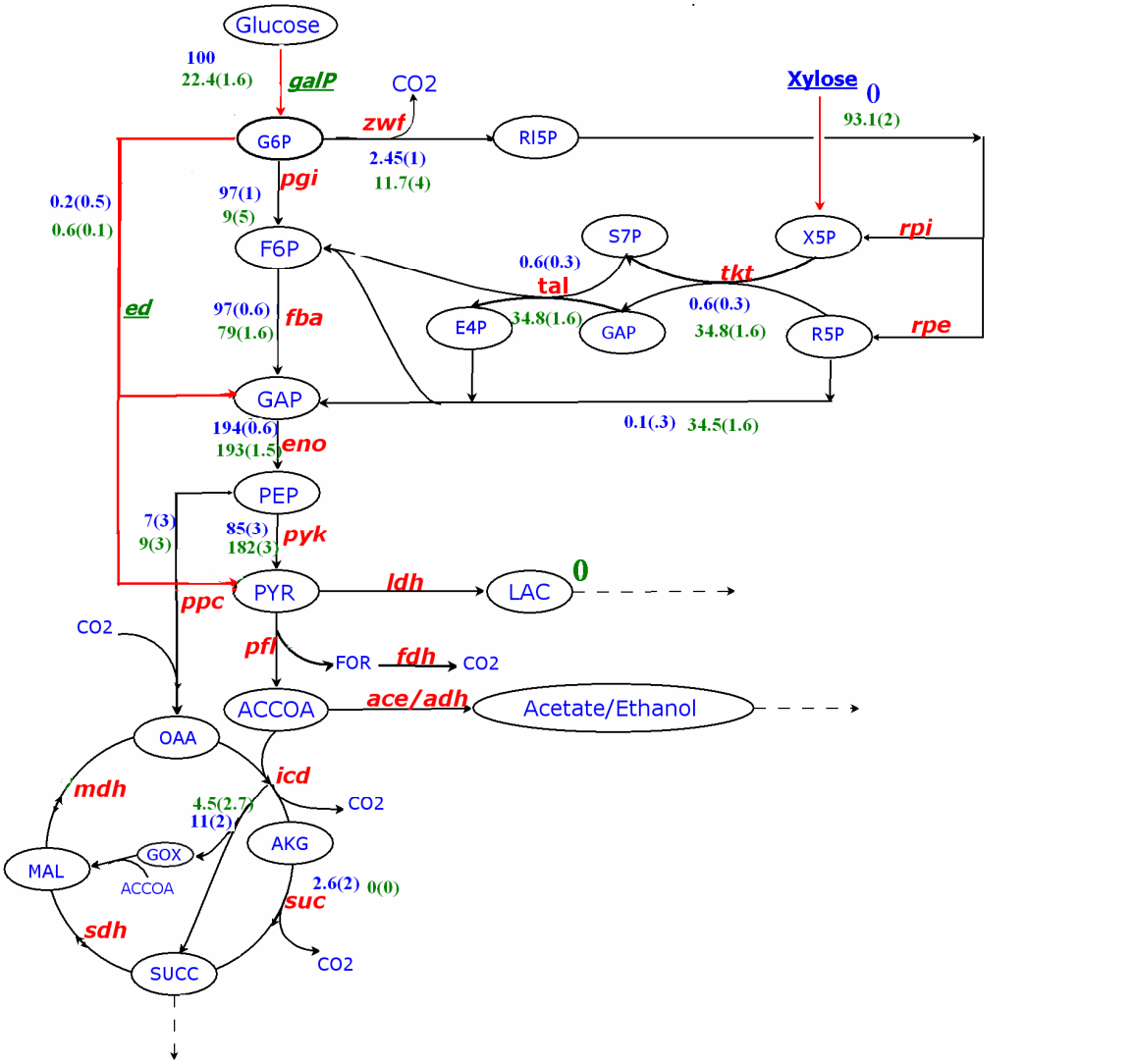

Figure 5.4: The metabolic flux map of wild and *ptsG* mutant of E. coli on sugar mixture. Flux values are reported relative to 600 moles of carbon consumed (100 moles of glucose in wild type).Values in blue represent fluxes in wild type while green values represent the ptsG mutant. Value in braces are standard deviation of the fluxes.

Conclusion

Metabolic flux analysis (MFA) is an important tool in metabolic engineering. In this work, ^{13}C MFA has been used to estimate metabolic fluxes in *Escherichia coli* under anaerobic condition. It was found that extracellular measurement plays an important role the estimation of metabolic fluxes even with ^{13}C labeling data. However, it is possible to estimate some of the intracellular fluxes even when some of the measurements are not used in the flux calculation. Additionally it is possible to estimate some of the extracellular fluxes using ^{13}C MFA.

In the second part of the thesis, comparative flux analysis was carried out in wild and *ptsG* mutant of *E. coli*. Its *ptsG* mutant was found to consume both glucose and xylose at the same time. However, it consumed glucose much slower than wild type. It also grew much slower than wild type. The metabolic flux analysis reveals slow growth in the *ptsG* is due to less ATP production. However, methylglyoxal synthesis could not be ruled out.

Future directions

The wild type consumed only glucose while *ptsG* mutant consumed both glucose and xylose. MFA with wild type and *ptsG* grown on xylose will be an interesting study to see the effect of coutilization of glucose and xylose. Additionally the comparison of MFA results with DNA microarray studies can be used find whether the changes in the flux level is correlated the change gene expression level. The integration of these using mathematical models might provide insight next set of genetic modification for the metabolic engineering of *E. coli* efficient in utilization of plant biomass.

6 Appendix

Appendix A2: Stiochiometric Matrix (G) of metabolic network of *E.coli* used in conventional metabolic flux analysis

Metabolite	Reactions																																
	In	pts	pgif	fbaf	enof	pykf	zwf	tktf	tkt2f	talf	ppcf	malf	frdf	pflf	ackf	cs	fdhf	Ldh	ac	for	succ	co2	lac	g6p	f6p	r5p	e4p	t3p1	pep	pyr	accoa	akg	oaa	serf	glyf	bser	bgly
GLU	1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PEP	0	-1	0	0	1	-1	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0
G6P	0	1	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0
PYR	0	1	0	0	0	1	0	0	0	0	0	0	0	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0
F6P	0	0	1	-1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0
CO2	0	0	0	0	0	0	1	0	0	0	-1	0	0	0	0	1	1	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
G3P	**0**	**0**	**0**	**2**	**-1**	**0**	**0**	**1**	**1**	**-1**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**0**	**-1**	**0**	**0**	**0**	**0**	**-1**	**0**	**0**	**0**	**0**
OAA	0	0	0	0	0	0	0	0	0	0	1	-1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0
SUCC	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ACCOA	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0
FOR	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	-1	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R5P	0	0	0	0	0	0	1	-2	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0
S7P	0	0	0	0	0	0	0	1	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
E4P	0	0	0	0	0	0	0	0	-1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0
AC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
AKG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0	0	0	0	0
LAC	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MAL	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
SER	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	-1	-1	0	0
GLY	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	-1

Appendix A2: Sentivity Matrix for Conventional Metabolic flux analysis.

Calculated Fluxes(Dependent Fluxes)	External Flux Measurements(Independent Fluxes)															
	glucose	*acetate*	*formate*	*succinate*	*lactate*	*g6p*	*f6p*	*r5p*	*e4p*	*t3p1*	*pep*	*pyr*	*accoa*	*akg*	*oaa*	*bser*
pts	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
pgi	5	-3	0	-3	-3	-5	-6	-5	-4	-3	-3	-3	-3	-6	-3	-3
fba	1	-1	0	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-2	-1	-1
eno	0	-1	0	-1	-1	0	0	0	0	0	-1	-1	-1	-2	-1	0
pyk	1	-1	0	0	-1	0	0	0	0	0	0	-1	-1	-1	0	0
zwf	**-6**	**3**	**0**	**3**	**3**	**6**	**6**	**5**	**4**	**3**	**3**	**3**	**3**	**6**	**3**	**3**
tktf	-2	1	0	1	1	2	2	2	1	1	1	1	1	2	1	1
tkt2f	-2	1	0	1	1	2	2	2	2	1	1	1	1	2	1	1
talf	-2	1	0	1	1	2	2	2	1	1	1	1	1	2	1	1
ppcf	0	0	0	-1	0	0	0	0	0	0	0	0	0	-1	-1	0
malf	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0	0
frdf	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
pflf	0	-1	0	0	0	0	0	0	0	0	0	0	-1	-1	0	0
ackf	0	-1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
cs	0	0	0	0	0	0	0	0	0	0	0	0	0	-1	0	0
fdhf	0	-1	1	0	0	0	0	0	0	0	0	0	-1	-1	0	0
ldh	0	0	0	0	-1	0	0	0	0	0	0	0	0	0	0	0
co2	-6	2	1	4	3	6	6	5	4	3	3	3	2	5	4	3
serf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	-1
glyf	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Table A3:Biomass flux in mol per 100 mol of glucose consumed[1]

Precursor	U-^{13}C	1-^{13}C
G6P	0.46	0.44
F6P	0.16	0.15
R5P	2.02	1.91
E4P	0.81	0.77
T3P[2]	1.88	1.79
PEP	1.17	1.11
PYR	6.37	6.04
ACCOA	8.43	7.99
AKG	2.43	2.30
OAA	4.02	3.81
Ser	0.46	0.44
Gly	1.31	1.24

[1] Calculated from Table A2 and A3 using biomas yield from Table 4.1

[2]T3P flux is calculated by adding 3PG and T3P flux and subtracting glycine serine flux as 3PG is precursor of serine and glycine

Table A4: External Fluxes used in flux calculations

	U-^{13}C		1-^{13}C	
Product	Average	S.D	Average	S.D
Glucose	100.00	0.00	100.00	0.00
Lactate	11.78	4.84	11.28	4.25
Succinate	5.49	0.81	5.22	1.08
Formate	136.13	4.07	136.32	5.01
Acetate	70.90	10.02	72.37	13.36
Ethanol	62.86	0.99	60.13	0.18
Acetate+ Ethanol	133.77	11.01	132.51	13.54

Appendix A5

Product Yield and Biomass Yield Calculations

If P_t and P_o denotes the product concentration at time=t and time=0, moles of product formed ΔP is

$$\Delta P = P_t - P_o$$

The moles of substrate ΔS consumed can be calculated similarly.

Product Yield (mol/mol) is defined as moles of product formed per mole of substrate consumed.
Hence,

$$Y_{p/s} = \frac{\Delta P}{\Delta S}.$$

Similarly,

Biomass Yield is calculated as $Y_{x/s} = \frac{\Delta X}{\Delta S}$

Where ΔX is the number of moles of carbon in the biomass which can be calculated from the OD by using 1 OD =0.36 g Dry Weight (DW) and 1gm of biomass contains 0.5 gm of carbon.

If steady state is assumed during exponential growth i.e. product yield and specific growth rate is constant. Hence, product yield can be calculated for different set of initial time and final time. Average and standard deviation of product and biomass yield was calculated from three set product yield for three different set of initial and final time.

Specific Substrate Uptake rate Calculations

Glucose uptake is related to the cell growth as

$$\frac{dS}{dt} = \frac{dX}{dt} \times \frac{1}{Y_{x/s}}$$

Hence, specific glucose uptake rate can be written as

$$\frac{dS}{Xdt} = \frac{dX}{Xdt} \times \frac{1}{Y_{x/s}}$$

Substituting $\mu = \frac{dX}{XdT}$, we get

$$\frac{dS}{Xdt} = \mu \times \frac{1}{Y_{x/s}}$$

Hence, specific glucose uptake (mol hr^{-1} per g of biomass) is calculated as from specific growth rate and biomass yield. Specific growth rate can be growth profile (OD vs time) and were found to be 0.57 hr^{-1} and 0.58 hr-1 in U-^{13}C and 1-^{13}C experiment respectively. Biomass yield was found to be 0.16 g of biomass/g of substrate and 0.15 g of biomass/g

of substrate. Specific glucose rate are 3.58 g of substrate/g biomass hr and 3.84 g of substrate/g biomass hr are for U-13C and 1-^{13}C experiment.

Appendix A6

Table A5: Fermentation data of U-^{13}C experiment in milimoles

Time(in hrs)	Biomass[1]	glucose	Lactate	Succinate	Formate	Acetate	Ethanol
0.00	0.78	55.51	0.03	0.02	0.06	0.03	0.04
1.68	0.99	55.65	0.00	0.06	0.87	0.45	0.44
2.85	1.77	54.91	0.01	0.05	2.05	1.30	0.73
3.77	3.21	53.98	0.03	0.25	4.37	2.11	1.20
5.02	6.47	49.92	0.36	0.42	8.92	5.15	4.04
5.60	9.56	46.74	1.25	0.49	12.89	6.45	5.94

[1]Moles carbon in biomass,calculated from OD using 1OD=0.36 g DW and 50% of biomass is carbon

Table A6: Fermentation data of 1-^{13}C experiment in millimoles

Time(in hrs)	Biomass[1]	glucose	Lactate	Succinate	Formate	Acetate	Ethanol
0.00	0.60	55.13	0.00	0.02	0.07	0.00	0.02
1.68	0.96	55.01	0.00	0.06	0.89	0.42	0.41
2.85	1.80	54.63	0.01	0.07	2.03	1.30	0.68
3.77	3.03	53.63	0.02	0.26	4.14	2.31	1.12
5.02	6.53	49.43	0.36	0.42	8.75	5.29	3.77
5.60	9.75	45.12	1.34	0.51	14.47	7.16	6.38

[1]Moles carbon in biomass, calculated from OD using 1OD=0.36 g DW and 50% of biomass is carbon

Cross peak	U-13C		1-13C			
(multiplet)	Intensity		Intensity	SD	Precursor	Isotopomer
Gly α (s)	0.19	0.20	0.15	0.19	Gly	[1**2**x]
Gly α (d)	0.82	0.80	0.85	0.81	Gly	[**12**x]
Ser β (s)	0.18	0.19	0.62	0.58	Ser	[x2**3**]
Ser β (d)	0.82	0.81	0.38	0.42	Ser	[x**23**]
Ser α (s)	0.13	0.12	0.14	0.11	Ser	[1**2**3]
Ser α (d1)	0.08	0.07	0.12	0.14	Ser	[**12**3]
Ser α (d2)	0.05	0.09	0.07	0.07	Ser	[1**23**]
Ser α (dd)	0.74	0.73	0.67	0.68	Ser	[**123**]
Ala α (s)	0.11	0.11	0.11	0.10	Pyr	[1**2**3]
Ala α (d1)	0.00	0.01	0.03	0.01	Pyr	[**12**3]
Ala α (d2)	0.48	0.46	0.42	0.43	Pyr	[1**23**]
Ala α (dd)	0.40	0.42	0.44	0.46	Pyr	[**123**]
Ala β (s)	0.12	0.12	0.56	0.55	Pyr	[x2**3**]
Ala β (d)	0.88	0.88	0.44	0.45	Pyr	[x**23**]
Ile γ2 (s)	0.11	0.12	0.56	0.55	Pyr	[x2**3**]
Ile γ2 (d)	0.89	0.88	0.44	0.45	Pyr	[x**23**]
Leu δ1 (s)	0.15	0.12	0.63	0.55	Pyr	[x2**3**]
Leu δ1 (d)	0.85	0.88	0.37	0.45	Pyr	[x**23**]
Val α (s)	0.49	0.51	0.43	0.47	Pyr	[1**2**x]·[x2x]
Val α (d1)	0.42	0.38	0.39	0.42	Pyr	[**12**x]·[x2x]
Val α (d2)	0.05	0.06	0.11	0.06	Pyr	[1**2**x]·[x**2**x]
Val α (dd)	0.03	0.05	0.08	0.05	Pyr	[**12**x]·[x**2**x]
Val γ1 (s)	0.13	0.12	0.59	0.55	Pyr	[x2**3**]
Val γ1 (d)	0.87	0.88	0.41	0.45	Pyr	[x**23**]
Val γ2 (s)	0.88	0.89	0.85	0.89	Pyr	[x2x]·[xx**3**]
Val γ2 (d)	0.12	0.11	0.15	0.11	Pyr	[x**2**x]·[xx**3**]
Leu δ2 (s)	0.91	0.89	0.90	0.89	Pyr	[x2x]·[xx**3**]
Leu δ2 (d)	0.09	0.11	0.10	0.11	Pyr	[x**2**x]·[xx**3**]
Phe α (s)	0.11	0.11	0.07	0.10	PEP	[1**2**3]
Phe α (d1)	0.01	0.01	0.00	0.01	PEP	[**12**3]
Phe α (d2)	0.09	0.12	0.10	0.08	PEP	[1**23**]
Phe α (dd)	0.80	0.76	0.82	0.80	PEP	[**123**]
Tyr β (s)	0.07	0.10	0.50	0.49	PEP	[x2**3**]·[x2x]
Tyr β (d)	0.79	0.80	0.49	0.46	PEP	[x**23**]·[x2x] + [x23]·[x2x]
Tyr β (t)	0.14	0.10	0.00	0.05	PEP	[x**23**]·[x**2**x]

Cross peak	U-13C		1-13C			
(multiplet)	Intensity		Intensity	SD	Precursor	Isotopomer
Leu α (s)	0.14	0.10	0.50	0.49	ACoA/Pyr	[1**2**]·[x2x]
Leu α (d1)	0.77	0.78	0.33	0.40	ACoA/Pyr	[**12**]·[x2x]
Leu α (d2)	0.04	0.01	0.12	0.06	ACoA/Pyr	[1**2**]·[x**2**x]
Leu α (dd)	0.05	0.10	0.05	0.05	ACoA/Pyr	[**12**]·[x**2**x]
Leu b (s)	0.81	0.79	0.67	0.69	ACoA/Pyr	[x2]·[x**2**x].[x2x]
Leu b (d)	0.16	0.20	0.29	0.29	ACoA/Pyr	[x**2**]·[x**2**x].[x2x]
Leu b (t)	0.03	0.01	0.05	0.03		+[x2]·[x2x].[x2x]
His δ2 (s)	0.25	0.28	0.68	0.68	R5P	[**1**2xxx]
His δ2 (d)	0.75	0.72	0.32	0.32	R5P	[**12**xxx]
His β (s)	0.12	0.10	0.08	0.06	P5P	[x2**3**4x]
His β (d1)	0.05	0.00	0.05	0.00	P5P	[x**23**4x]
His β (d2)	0.01	0.06	0.14	0.01	P5P	[x2**34**x]
His β (dd)	0.82	0.83	0.73	0.92	P5P	[x**234**x]
Tyr δ1 (s)	0.17	0.10	0.55	0.01	PEP/E4P	[x2**3**]·[1xxx]
Tyr δ1 (d)	0.75	0.80	0.41	0.01	PEP/E4P	[x**23**]·[1xxx] + [x2**3**]·[**1**xxx]
Tyr δ1 (t)	0.07	0.10	0.04	0.01	PEP/E4P	[x**23**]·[**1**xxx]
Arg β (s)	0.10	0.10	0.09	0.13	AKG	[x2**3**4x]
Arg β (d)	0.80	0.80	0.68	0.68	AKG	[x**23**4x] + [x2**34**x]
Arg β (t)	0.10	0.10	0.24	0.19	AKG	[x**234**x]
Arg δ (s)	0.15	0.12	0.17	0.11	AKG	[xxx4**5**]
Arg δ (d)	0.85	0.88	0.83	0.89	AKG	[xxx**45**]
Pro α (s)	0.10	0.11	0.51	0.47	AKG	[xx3**45**]
Pro α (d1)	0.01	0.01	0.09	0.06	AKG	[xx**345**]
Pro α (d2)	0.73	0.76	0.35	0.39	AKG	[xx3**45**]
Pro α (dd)	0.18	0.12	0.06	0.08	AKG	[xx**345**]
Pro γ (s)	0.08	0.10	0.48	0.48	AKG	[x2**3**4x]
Pro γ (d)	0.81	0.80	0.43	0.46	AKG	[x**23**4x] + [x2**34**x]
Pro γ (t)	0.11	0.10	0.09	0.05	AKG	[x**234**x]
Asp α (s)	0.13	0.11	0.16	0.15	OAA	[1**2**3x]
Asp α (d1)	0.01	0.01	0.05	0.02	OAA	[**12**3x]
Asp α (d2)	0.19	0.15	0.13	0.13	OAA	[1**23**x]
Asp α (dd)	0.67	0.73	0.67	0.71	OAA	[**123**x]
Asp β (s)	0.57	0.11	0.12	0.13	OAA	[x2**3**4]
Asp β (d1)	0.38	0.76	0.64	0.68	OAA	[x**23**4]
Asp β (d2)	0.03	0.01	0.25	0.19	OAA	[x2**34**]
Asp β (dd)	0.03	0.12	0.46	0.48	OAA	[x**234**]

Cross peak	U-13C		1-13C			
(multiplet)	Intensity		Intensity	SD	Precursor	Isotopomer
Thr γ2 (s)	0.89	0.86	0.73	0.73	OAA	[xx3**4**]
Thr γ2 (d)	0.11	0.14	0.27	0.27	OAA	[xx**34**]
Ile α (s)	0.23	0.23	0.21	0.25	OAA/Pyr	[1**2**xx]·[x2x]
Ile α (d1)	0.61	0.66	0.62	0.64	OAA/Pyr	[**12**xx]·[x2x]
Ile α (d2)	0.04	0.03	0.08	0.03	OAA/Pyr	[1**2**xx]·[x**2**x]
Ile α (dd)	0.12	0.08	0.09	0.08	OAA/Pyr	[**12**xx]·[x**2**x]
Ile γ1(s)	0.80	0.77	0.78	0.76	Pyr/OAA	[x2x]·[xx**3**4]
Ile γ1(d)	0.19	0.21	0.21	0.22	Pyr/OAA	[x**2**x]·[xx**3**4] + [x2x]·[xx**34**]
Ile γ1(t)	0.00	0.01	0.00	0.02	Pyr/OAA	[x**2**x]·[xx**34**]
Ile δ (s)	0.90	0.86	0.78	0.73	OAA	[xx3**4**]
Ile δ (d)	0.10	0.14	0.22	0.27	OAA	[xx**34**]
Lys β (s)	0.10	0.11	0.47	0.48	OAA/Pyr	½{[x2**3**4] + [x2**3**]·[xxx4]}
Lys β (d)	0.80	0.79	0.47	0.46	OAA/Pyr	½{[x**23**4] + [x2**34**] + [x**23**]·[xxx4] + [x2**3**]·[xxx**4**]}
Lys β (t)	0.10	0.11	0.07	0.06	OAA/Pyr	½{[x**234**] + [x**23**]·[xxx**4**]}
Lys γ (s)	0.79	0.77	0.56	0.57	OAA/Pyr	[xx3**4**]·[xx3]
Lys γ (d)	0.18	0.22	0.37	0.37	OAA/Pyr	[xx**34**]·[xx3] + [xx3**4**]·[xx**3**]
Lys γ (t)	0.03	0.01	0.07	0.06	OAA/Pyr	[xx**34**]·[xx**3**]
Lys δ (s)	0.10	0.11	0.51	0.48	OAA/Pyr	½{[x2**3**4] + [x2**3**]·[xxx4]}
Lys δ (d)	0.78	0.79	0.42	0.46	OAA/Pyr	½{[x**23**4] + [x2**34**] + [x**23**]·[xxx4] + [x2**3**]·[xxx**4**]}
Lys δ (t)	0.12	0.11	0.07	0.06	OAA/Pyr	½{[x**234**] + [x**23**]·[xxx**4**]}

AppendixA8: Flux estmatied by 13C MFA and concention MFA for 1-labeled expeirment and U-13C experiment.Fluxes are reported for conventional flux analysis.

Reaction *Name*	1NM		UNM		1edmez		uedmez		UTGnm		1TGnm		1-TG		U-TG		ucm		1cm		CMFA_IRR		CMFA	
	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD
PTS tranport of Glucose																								
In	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00
pts	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00	100.00	0.00
Glycolysis																								
pgif	96.17	1.27	77.11	10.89	96.98	1.65	36.67	9.73	64.24	9.30	97.58	0.93	98.42	0.60	63.10	9.26	75.65	13.54	96.92	1.25	55.14	33.71	61.63	21.95
pgib	0.80	0.35	0.58	0.34	0.70	0.43	0.80	0.14	0.67	0.23	0.75	0.38	0.69	0.44	0.67	0.26	0.62	0.34	0.85	0.30	NA	NA	NA	NA
fbaf	96.88	0.45	90.30	3.64	96.21	1.55	53.68	8.61	86.00	3.10	97.37	0.31	97.61	0.22	85.62	3.09	89.80	4.51	97.15	0.42	82.96	11.24	85.13	7.32
enof	192.04	0.57	183.17	3.63	191.19	1.67	146.57	8.61	178.89	3.10	192.59	0.45	192.69	0.49	178.51	3.09	182.70	4.51	192.35	0.50	175.86	11.24	178.02	7.32
enob	0.86	0.24	0.23	0.20	0.94	0.12	0.30	0.34	0.18	0.19	0.91	0.20	0.94	0.13	0.19	0.19	0.19	0.22	0.91	0.16	NA	NA	NA	NA
pykf	82.10	1.54	72.21	5.09	17.81	21.21	1.82	7.34	65.80	3.15	78.04	1.19	78.17	1.57	65.52	3.26	71.58	6.24	82.80	1.69	62.78	11.24	64.87	7.34
pykb	0.18	0.08	0.08	0.12	0.23	0.24	0.28	0.27	0.22	0.05	0.11	0.06	0.16	0.05	0.21	0.04	0.12	0.16	0.20	0.09	NA	NA	NA	NA
Pentose Pathway																								
zwf	3.42	1.25	22.40	10.91	1.32	0.74	28.17	8.87	35.30	9.30	2.00	0.88	1.15	0.59	36.44	9.26	23.89	13.54	2.68	1.25	44.40	33.71	37.91	21.95
tktf	0.77	0.42	7.02	3.67	0.08	0.29	8.99	2.96	11.36	3.11	0.33	0.29	0.06	0.26	11.74	3.09	7.56	4.51	0.56	0.41	14.40	11.24	12.23	7.32
tktb	0.86	0.28	0.74	0.30	0.98	0.03	0.82	0.23	0.63	0.23	0.98	0.04	0.98	0.10	0.60	0.23	0.76	0.23	0.83	0.34	NA	NA	NA	NA
tkt2f	0.01	0.44	6.25	3.65	-0.70	0.26	8.18	2.96	10.55	3.10	-0.46	0.31	-0.74	0.22	10.93	3.09	6.75	4.51	-0.19	0.44	13.58	11.24	11.42	7.32
tkt2b	0.88	0.22	0.22	0.42	0.47	0.36	0.33	0.29	0.34	0.26	0.76	0.26	0.45	0.37	0.35	0.24	0.27	0.28	0.82	0.26	NA	NA	NA	NA
talf	0.76	0.43	7.03	3.67	0.08	0.29	8.99	2.96	11.36	3.11	0.33	0.29	0.06	0.26	11.74	3.09	7.56	4.51	0.56	0.41	14.40	11.24	12.23	7.32
talb	0.86	0.08	0.44	0.30	0.87	0.07	0.39	0.27	0.56	0.29	0.88	0.06	0.88	0.12	0.57	0.31	0.48	0.30	0.88	0.08	NA	NA	NA	NA
Anapletoric Reactions and TCA cycle																								
ppcf	7.17	1.61	9.74	1.86	70.62	20.93	43.59	9.91	11.92	0.68	11.83	0.98	11.87	1.20	11.82	0.82	9.95	2.00	6.92	1.53	11.92	0.76	11.99	0.85
ppcb	0.93	0.14	0.91	0.24	0.65	0.11	0.28	0.16	0.37	0.23	0.91	0.12	0.93	0.07	0.40	0.22	0.98	0.01	0.96	0.01	NA	NA	NA	NA
malf	-5.41	1.82	-0.23	2.68	53.75	20.86	36.37	9.97	5.47	0.68	5.77	0.90	5.58	0.98	5.38	0.82	0.59	3.25	-5.22	1.85	5.47	0.76	5.54	0.85
malr	0.69	0.93	0.76	1.02	0.94	0.05	0.52	0.15	0.48	0.20	0.88	0.12	0.90	0.08	0.51	0.21	0.96	0.10	0.82	0.21	NA	NA	NA	NA
frdf	1.46	1.15	-1.91	2.03	4.67	1.49	-5.00	1.14	-5.47	0.68	-5.80	0.91	-5.59	0.99	-5.38	0.82	-2.94	2.06	1.12	1.09	-5.47	0.76	-5.54	0.85
frds	0.45	0.34	0.75	0.24	0.62	0.35	0.78	0.20	0.58	0.25	0.67	0.22	0.67	0.20	0.58	0.24	0.72	0.28	0.49	0.34	NA	NA	NA	NA
frdb	0.84	0.22	0.92	0.11	0.23	0.28	0.72	0.24	0.33	2.26	0.76	0.99	0.85	0.64	0.56	0.24	0.87	0.21	0.88	0.22	NA	NA	NA	NA
pflf	165.02	4.12	153.79	7.90	165.48	6.75	148.56	6.77	147.61	5.66	160.95	3.91	160.56	6.67	147.42	7.05	151.22	8.32	167.55	6.32	144.79	10.51	147.06	6.32
pflb	0.51	0.09	0.50	0.15	0.54	0.01	0.57	0.03	0.56	0.06	0.51	0.07	0.52	0.04	0.56	0.01	0.55	0.03	0.52	0.02	NA	NA	NA	NA
ackf	144.41	4.39	137.32	6.53	136.45	7.74	136.70	6.64	136.76	5.66	150.80	4.00	149.16	6.92	136.57	7.05	135.12	7.19	146.31	6.86	133.93	10.51	136.21	6.32
cs	4.76	1.94	3.88	1.71	5.36	2.29	2.97	0.89	2.43	0.00	2.31	0.21	2.52	0.20	2.43	0.00	2.99	0.96	4.22	1.55	2.43	0.00	2.43	0.00
tca	2.38	1.93	1.45	1.71	2.91	2.28	0.55	0.89	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.57	0.96	1.70	1.52	NA	NA	NA	NA
gos	3.91	1.58	2.10	1.38	7.57	1.61	0.23	0.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.34	1.48	4.11	1.42	NA	NA	NA	NA
gos1	3.91	1.58	2.10	1.38	7.57	1.61	0.23	0.34	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	2.34	1.48	4.11	1.42	NA	NA	NA	NA
Fermentative reactions																								
fdhf	86.56	44.80	101.28	53.85	28.28	12.41	13.05	11.72	83.60	46.08	117.60	51.86	26.46	12.04	11.91	11.35	16.49	13.13	33.65	12.47	8.91	11.21	10.99	6.52
Ldh	11.27	3.98	12.03	4.93	13.30	5.03	13.18	5.77	11.83	4.64	11.11	3.89	11.53	5.18	11.73	5.58	13.99	5.28	9.38	4.92	11.62	4.62	11.44	4.58
ac	144.41	4.39	137.32	6.53	136.45	7.74	136.70	6.64	136.76	5.66	150.80	4.00	149.16	6.92	136.57	7.05	135.12	7.19	146.31	6.86	133.93	10.51	136.21	6.32
for	78.47	45.59	52.54	53.94	137.19	12.76	135.51	10.84	64.01	46.34	43.36	51.34	134.10	12.61	135.52	10.63	134.73	11.34	133.90	12.37	135.88	3.80	136.07	3.94
succ	4.82	1.05	5.46	1.09	5.82	1.10	5.78	0.89	5.48	0.68	5.79	0.90	5.59	0.98	5.38	0.82	5.85	0.97	4.70	1.06	5.47	0.76	5.54	0.85
co2	91.13	44.79	120.52	54.77	34.43	13.22	34.05	13.54	110.71	45.73	111.27	52.34	19.36	12.39	40.26	14.11	35.30	13.88	36.44	13.29	45.13	24.86	40.65	18.54
lac	11.27	3.98	12.03	4.93	13.30	5.03	13.18	5.77	11.83	4.64	11.11	3.89	11.53	5.18	11.73	5.58	13.99	5.28	9.38	4.92	11.62	4.62	11.44	4.58
Biosynthetic Reactions																								
g6p	0.39	0.04	0.46	0.00	0.43	0.06	0.46	0.00	0.46	0.00	0.40	0.04	0.42	0.06	0.46	0.00	0.46	0.00	0.41	0.05	0.46	0.00	0.46	0.00
f6p	0.13	0.01	0.16	0.00	0.15	0.02	0.16	0.00	0.16	0.00	0.14	0.02	0.15	0.02	0.16	0.00	0.16	0.00	0.14	0.02	0.16	0.00	0.16	0.00
r5p	1.80	0.15	2.02	0.00	1.86	0.20	2.02	0.00	2.02	0.00	1.75	0.14	1.76	0.17	2.02	0.00	2.02	0.00	1.74	0.13	2.02	0.00	2.02	0.00
e4p	0.75	0.07	0.81	0.00	0.78	0.08	0.81	0.00	0.81	0.00	0.79	0.08	0.80	0.09	0.81	0.00	0.81	0.00	0.75	0.08	0.81	0.00	0.81	0.00
t3p1	1.74	0.24	1.88	0.00	1.79	0.24	1.88	0.00	1.88	0.00	1.68	0.21	1.80	0.24	1.88	0.00	1.88	0.00	1.76	0.24	1.88	0.00	1.88	0.00
pep	1.11	0.15	1.17	0.00	1.14	0.15	1.17	0.00	1.17	0.00	1.08	0.15	1.13	0.15	1.17	0.00	1.17	0.00	1.12	0.15	1.17	0.00	1.17	0.00
pyr	5.80	0.55	6.37	0.00	6.29	0.66	6.37	0.00	6.37	0.00	5.97	0.44	6.07	0.65	6.37	0.00	6.37	0.00	5.87	0.68	6.37	0.00	6.37	0.00
accoa	8.04	0.76	8.43	0.00	8.51	1.00	8.43	0.00	8.43	0.00	7.86	0.65	8.88	0.70	8.43	0.00	8.43	0.00	8.80	0.80	8.43	0.00	8.43	0.00
akg	2.38	0.22	2.43	0.00	2.45	0.28	2.43	0.00	2.43	0.00	2.31	0.21	2.52	0.20	2.43	0.00	2.43	0.00	2.52	0.21	2.43	0.00	2.43	0.00
oaa	3.91	0.32	4.02	0.00	3.94	0.44	4.02	0.00	4.02	0.00	3.76	0.29	3.76	0.37	4.02	0.00	4.02	0.00	3.81	0.38	4.02	0.00	4.02	0.00
serf	1.66	0.11	1.77	0.00	1.62	0.18	1.77	0.00	1.77	0.00	1.63	0.12	1.52	0.11	1.77	0.00	1.77	0.00	1.51	0.12	1.77	0.00	1.77	0.00
glyf	1.21	0.12	1.30	0.03	1.19	0.16	1.31	0.00	1.31	0.01	1.20	0.12	1.10	0.10	1.31	0.00	1.31	0.00	1.10	0.10	1.31	0.00	1.31	0.00
glyb	0.21	0.02	0.13	0.02	0.23	0.02	0.10	0.01	0.12	0.01	0.23	0.02	0.23	0.02	0.12	0.01	0.13	0.01	0.21	0.02	0.53	0.30	0.51	0.31
bser	0.44	0.03	0.46	0.00	0.43	0.04	0.46	0.00	0.46	0.00	0.43	0.03	0.42	0.04	0.46	0.00	0.46	0.00	0.41	0.04	NA	NA	NA	NA
bgly	1.22	0.11	1.31	0.00	1.19	0.16	1.31	0.00	1.31	0.00	1.21	0.11	1.10	0.09	1.31	0.00	1.31	0.00	1.10	0.10	1.31	0.00	1.31	0.00
ed					1.27	1.53	34.70	9.49																
mezf					65.99	21.18	31.59	9.98																
mezb					0.46	0.11	0.76	0.16																
chi-sq	**651**		**647**		**544.5**		**467**		**681**		**728**		**729.5**		**681**		**646**		**652**					

Appendix B1: Relative metabolic fluxes of wild and ptsG mutant per 600 moles of glucose consumed. Fluxes are reported for all the replciates and wild type and ptsG flux are estimated as the averae of replicates.

Reaction	w3110						ptsG				W3110		ptsG	
Name	D		E		F		E		F		Avg of DEF		Avg EF	
	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD	Average	SD
PTS tranport of Glucose														
glu	100.000	0.000	100.000	0.000	100.000	0.000	24.900	0.000	21.370	0.000	100.000	0.000	22.404	1.609
xyl	0.000	0.000	0.000	0.000	0.000	0.000	90.110	0.000	94.350	0.000	0.000	0.000	93.108	1.932
NaHCO3	1.072	2.117	0.540	1.138	0.069	0.237	67.811	10.043	145.815	13.485	0.667	1.600	122.968	37.705
pts	100.000	0.000	100.000	0.000	100.000	0.000	24.900	0.000	21.370	0.000	100.000	0.000	22.404	1.609
Glycolysis														
pgif	97.646	1.279	96.452	0.489	96.695	0.756	14.153	6.542	8.160	1.724	97.026	1.107	9.915	4.691
pgib	0.446	0.421	0.332	0.335	0.806	0.350	0.240	0.315	0.135	0.186	0.487	0.417	0.166	0.236
fbaf	97.835	0.650	97.484	0.386	97.348	0.706	80.093	2.465	78.729	0.944	97.607	0.622	79.129	1.668
enof	**194.734**	**0.650**	**194.324**	**0.386**	**194.111**	**0.706**	**193.198**	**2.465**	**192.496**	**0.944**	**194.455**	**0.641**	**192.702**	**1.581**
enob	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
pykf	88.428	1.521	82.894	1.016	82.323	1.030	185.164	2.637	181.520	2.556	85.179	3.131	182.587	3.065
pykb	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
edf	**0.253**	**0.531**	**0.148**	**0.375**	**0.429**	**0.698**	**0.831**	**1.238**	**0.562**	**0.892**	**0.256**	**0.537**	**0.641**	**1.011**
Pentose Pathway														
zwf	1.407	0.453	3.122	0.272	2.602	0.306	10.211	6.314	12.685	1.718	2.446	0.996	11.681	3.950
tktf	0.235	0.151	0.802	0.091	0.623	0.102	33.297	2.105	35.531	0.573	0.578	0.331	34.783	1.637
tktb	0.963	0.033	0.156	0.225	0.989	0.002	0.541	0.129	0.201	0.102	0.646	0.423	0.302	0.191
tkt2f	-0.237	0.151	0.321	0.091	0.131	0.102	33.007	2.105	35.234	0.573	0.098	0.328	34.488	1.635
tkt2b	0.222	0.240	0.986	0.004	0.974	0.024	0.038	0.055	0.371	0.040	0.693	0.399	0.276	0.159
talf	0.235	0.151	0.802	0.091	0.623	0.102	33.297	2.105	35.531	0.573	0.578	0.331	34.783	1.637
talb	0.899	0.026	0.988	0.002	0.988	0.003	0.306	0.253	0.928	0.093	0.954	0.047	0.753	0.323
Anapletoric Reactions and TCA cycle														
ppcf	4.598	1.127	9.690	0.921	10.005	0.839	6.985	2.938	9.902	2.433	7.541	2.777	9.047	2.908
ppcb	0.986	0.005	0.946	0.013	0.932	0.020	0.926	0.184	0.931	0.162	0.960	0.026	0.929	0.169
malf	-13.684	1.563	-7.502	1.611	-4.790	1.271	-1.283	5.877	3.045	4.866	-9.595	4.028	1.778	5.537
malr	0.641	0.363	0.585	0.325	0.235	0.268	0.731	0.339	0.631	0.351	0.532	0.367	0.660	0.350
frdf	0.126	0.848	-1.035	1.860	-3.675	0.845	-4.683	2.938	-7.543	2.433	-1.114	1.944	-6.705	2.896
frds	0.369	0.430	0.272	0.360	0.347	0.437	0.431	0.318	0.353	0.365	0.331	0.411	0.376	0.353
frdb	0.328	0.304	0.205	0.247	0.162	0.171	0.454	0.395	0.392	0.328	0.249	0.269	0.410	0.350
pflf	184.979	1.474	179.271	0.935	178.888	0.861	183.721	2.505	179.750	2.518	181.673	3.143	180.913	3.095
pflb	0.567	0.010	0.532	0.007	0.498	0.008	0.494	0.013	0.521	0.011	0.540	0.028	0.513	0.017
ackf	150.575	0.841	150.931	1.869	152.952	0.831	167.912	4.495	166.783	2.475	151.223	1.590	167.113	3.235
ackb	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
cs	2.388	1.240	6.276	2.453	3.892	1.104	0.866	0.000	0.888	0.000	4.054	2.430	0.882	0.010
tca	0.978	1.240	4.839	2.453	2.420	1.104	0.000	0.000	0.000	0.000	2.621	2.423	0.000	0.000
gos	**13.558**	**1.127**	**8.536**	**0.921**	**8.465**	**0.839**	**5.967**	**2.938**	**4.497**	**2.433**	**10.709**	**2.698**	**4.928**	**2.672**
gos1	13.558	1.127	8.536	0.921	8.465	0.839	5.967	2.938	4.497	2.433	10.709	2.698	4.928	2.672
Fermentative reactions														
fdhf	90.252	12.619	77.075	9.208	0.878	1.280	116.283	5.647	119.592	8.151	65.944	36.581	118.623	7.646
Ldh	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ac	150.575	0.841	150.931	1.869	152.952	0.831	167.912	4.495	166.783	2.475	151.223	1.590	167.113	3.235
for	94.728	13.804	102.196	9.004	178.010	1.579	67.438	4.959	60.159	7.459	115.729	35.028	62.291	7.580
succ	14.410	0.000	14.410	0.000	14.560	0.000	10.650	0.000	12.040	0.000	14.443	0.062	11.633	0.634
co2	92.685	14.590	82.942	9.480	0.644	1.844	188.196	18.048	269.352	18.905	68.964	38.368	245.582	41.416
lac	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Biosynthetic Reactions														
g6p	0.268	0.000	0.273	0.000	0.280	0.000	0.165	0.000	0.169	0.000	0.272	0.005	0.167	0.002
f6p	0.093	0.000	0.094	0.000	0.097	0.000	0.057	0.000	0.058	0.000	0.094	0.002	0.058	0.001
r5p	1.174	0.000	1.196	0.000	1.225	0.000	0.721	0.000	0.739	0.000	1.193	0.020	0.734	0.008
e4p	0.472	0.000	0.481	0.000	0.493	0.000	0.290	0.000	0.297	0.000	0.480	0.008	0.295	0.003
t3p1	1.096	0.000	1.116	0.000	1.144	0.000	0.673	0.000	0.690	0.000	1.113	0.019	0.685	0.008
pep	0.679	0.000	0.691	0.000	0.708	0.000	0.417	0.000	0.427	0.000	0.690	0.011	0.424	0.005
pyr	3.702	0.000	3.772	0.000	3.864	0.000	2.274	0.000	2.331	0.000	3.762	0.063	2.315	0.026
accoa	4.899	0.000	4.992	0.000	5.114	0.000	3.010	0.000	3.085	0.000	4.979	0.083	3.063	0.034
akg	1.410	0.000	1.437	0.000	1.472	0.000	0.866	0.000	0.888	0.000	1.433	0.024	0.882	0.010
oaa	2.336	0.000	2.380	0.000	2.438	0.000	1.435	0.000	1.471	0.000	2.373	0.039	1.460	0.016
serf	1.029	0.000	1.048	0.000	1.074	0.000	0.632	0.000	0.648	0.000	1.045	0.017	0.643	0.007
glyf	0.761	0.000	0.775	0.000	0.794	0.000	0.467	0.000	0.479	0.000	0.773	0.013	0.476	0.005
glyb	0.658	0.023	0.668	0.019	0.948	0.007	0.638	0.021	0.846	0.009	0.725	0.120	0.785	0.096
bser	0.268	0.000	0.273	0.000	0.280	0.000	0.165	0.000	0.169	0.000	0.272	0.005	0.167	0.002
bgly	0.761	0.000	0.775	0.000	0.794	0.000	0.467	0.000	0.479	0.000	0.773	0.013	0.476	0.005

TableB2: Relative multiplet intensities of amino acids from protein hydrolysates with their standard deviations (SDs) from 2-D [13C, 1H] HSQC spectrum. s indicates singlet, d1 and d2 indicate the first and second doublet and dd indicates the double doublet. Bold faced carbon atom in Isotopomer indicates labeled carbon atom, normal font indicates unlabeled carbon atom and x indicates unknown labeling state of the carbon atom.

Cross peak	W3110-D		W3110-E		W3110-F		ptsG-E		ptsG-F			
(multiplet)	Intensity	SD	Intensity	SD	Intensity	SD	Intensity	SD	Intensity	SD	Precursor	Isotopomer
Gly α (s)	0.21	0.01	0.22	0.01	0.21	0.01	0.28	0.01	0.32	0.01	Gly	[**12**x]
Gly α (d)	0.79	0.01	0.78	0.01	0.79	0.01	0.72	0.01	0.67	0.01	Gly	[**12**x]
Ser β (s)	0.79	0.01	0.70	0.01	0.70	0.01	0.49	0.01	0.63	0.01	Ser	[x2**3**]
Ser β (d)	0.21	0.01	0.30	0.01	0.30	0.01	0.51	0.01	0.37	0.01	Ser	[x**23**]
Ala α (s)	0.10	0.01	0.08	0.01	0.09	0.01	0.03	0.01	0.06	0.01	Pyr	[12**3**]
Ala α (d1)	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.02	0.01	Pyr	[**12**3]
Ala α (d2)	0.43	0.01	0.45	0.01	0.45	0.01	0.45	0.01	0.51	0.01	Pyr	[1**23**]
Ala α (dd)	0.46	0.01	0.46	0.01	0.45	0.01	0.51	0.01	0.40	0.01	Pyr	[**123**]
Ala β (s)	0.60	0.01	0.64	0.00	0.58	0.01	0.25	0.01	0.27	0.01	Pyr	[x2**3**]
Ala β (d)	0.39	0.01	0.36	SD	0.42	0.01	0.75	0.01	0.73	0.01	Pyr	[x**23**]
Ile γ2 (s)	0.55	0.01	0.55	0.01	0.59	0.01	0.23	0.01	0.24	0.01	Pyr	[x2**3**]
Ile γ2 (d)	0.45	0.01	0.45	0.01	0.41	0.01	0.77	0.01	0.76	0.01	Pyr	[x**23**]
Val α (s)	0.48	0.01	0.55	0.01	0.54	0.01	0.45	0.01	0.40	0.01	Pyr	[12x]·[x2x]
Val α (d1)	0.44	0.01	0.38	0.01	0.36	0.01	0.36	0.01	0.37	0.01	Pyr	[**12**x]·[x2x]
Val α (d2)	0.05	0.01	0.04	0.01	0.06	0.01	0.12	0.01	0.12	0.01	Pyr	[12x]·[x**2**x]
Val α (dd)	0.03	0.01	0.03	0.01	0.04	0.01	0.07	0.01	0.11	0.01	Pyr	[**12**x]·[x**2**x]
Val γ1 (s)	0.56	0.01	0.58	0.01	0.60	0.01	0.23	0.01	0.24	0.01	Pyr	[x2**3**]
Val γ1 (d)	0.44	0.01	0.42	0.01	0.41	0.01	0.77	0.01	0.76	0.01	Pyr	[x**23**]
Val γ2 (s)	0.85	0.01	0.87	0.01	0.84	0.01	0.81	0.01	0.81	0.01	Pyr	[x2x]·[xx**3**]
Val γ2 (d)	0.16	0.01	0.13	0.01	0.16	0.01	0.19	0.01	0.19	0.01	Pyr	[x**2**x]·[xx**3**]
Leu δ2 (s)	0.93	0.01	0.90	0.01	0.90	0.01	0.81	0.01	0.82	0.01	Pyr	[x2x]·[xx**3**]
Leu δ2 (d)	0.07	0.01	0.09	0.01	0.10	0.01	0.19	0.01	0.18	0.01	Pyr	[x**2**x]·[xx**3**]
Phe α (s)	0.11	0.01	0.09	0.01	0.09	0.01	0.04	0.01	0.07	0.01	PEP	[12**3**]
Phe α (d1)	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	PEP	[**12**3]
Phe α (d2)	0.10	0.01	0.06	0.01	0.07	0.01	0.21	0.01	0.23	0.01	PEP	[1**23**]
Phe α (dd)	0.77	0.01	0.84	0.01	0.83	0.01	0.74	0.01	0.70	0.01	PEP	[**123**]
Phe β (s)	0.50	0.01	0.51	0.01	0.48	0.01	0.2	0.01	0.21	0.01	PEP	[x2**3**]·[x2x]
Phe β (d1)	0.41	0.01	0.42	0.01	0.37	0.01	0.7	0.01	0.64	0.01	PEP	[x**23**]·[x2x]
Phe β (d2)	0.05	0.01	0.05	0.01	0.06	0.01	0	0.01	0.04	0.01	PEP	[x2**3**]·[x**2**x]
Phe β (dd)	0.04	0.01	0.02	0.01	0.08	0.01	0.09	0.01	0.11	0.01	PEP	[x**23**]·[x**2**x]
Tyr α (s)	0.11	0.01	0.09	0.01	0.08	0.01	0.04	0.01	0.06	0.01	PEP	[12**3**]
Tyr α (d1)	0.08	0.01	0.09	0.01	0.09	0.01	0	0.01	0.00	0.01	PEP	[**12**3]
Tyr α (d2)	0.00	0.01	0.01	0.01	0.00	0.01	0.22	0.01	0.22	0.01	PEP	[1**23**]
Tyr α (dd)	0.80	0.01	0.82	0.01	0.83	0.01	0.74	0.01	0.72	0.01	PEP	[**123**]
Leu α (s)	0.46	0.01	0.55	0.01	0.54	0.01	0.21	0.01	0.15	0.01	ACoA/Pyr	[**12**]·[x2x]
Leu α (d1)	0.42	0.01	0.39	0.01	0.38	0.01	0.64	0.01	0.66	0.01	ACoA/Pyr	[**12**]·[x2x]
Leu α (d2)	0.05	0.01	0.05	0.01	0.06	0.01	0.04	0.01	0.02	0.01	ACoA/Pyr	[**12**]·[x**2**x]
Leu α (dd)	0.07	0.01	0.02	0.01	0.02	0.01	0.12	0.01	0.17	0.01	ACoA/Pyr	[**12**]·[x**2**x]
Leu b (s)	0.68	0.01	0.68	0.01	0.67	0.01	0.6	0.01	0.59	0.01	ACoA/Pyr	[x2]·[x**2**x].[x2x]
Leu b (d)	0.26	0.01	0.26	0.01	0.26	0.01	0.35	0.01	0.31	0.01	ACoA/Pyr	[x**2**]·[x**2**x].[x2x]
Leu b (t)	0.05	0.01	0.05	0.01	0.07	0.01	0.06	0.01	0.10	0.01		+[x2]·[x2x].[x2x]
His δ2 (s)	0.73	0.01	0.66	0.01	0.63	0.01					R5P	[**12**xxx]
His δ2 (d)	0.27	0.01	0.34	0.01	0.37	0.01					R5P	[**12**xxx]
Tyr δ1 (s)	0.55	0.01	0.55	0.01	0.57	0.01	0.22	0.01	0.19	0.01	PEP/E4P	[x2**3**]·[1xxx]
Tyr δ1 (d)	0.40	0.01	0.39	0.01	0.39	0.01	0.69	0.01	0.71	0.01	PEP/E4P	[x**23**]·[1xxx] + [x2**3**]·[**1**xxx]
Tyr δ1 (t)	0.05	0.01	0.04	0.01	0.04	0.01	0.1	0.01	0.10	0.01	PEP/E4P	[x**23**]·[**1**xxx]
Tyr ε1 (s)	0.35	0.01	0.22	0.01	0.22	0.01	0.13	0.01	0.23	0.01	PEP/E4P	[xx3]·[**12**xx]
Tyr ε1 (d)	0.25	0.01	0.30	0.01	0.23	0.01	0.27	0.01	0.32	0.01	PEP/E4P	[xx**3**]·[**12**xx] + [xx3]·[**12**xx]
Tyr ε1 (t)	0.40	0.01	0.47	0.01	0.56	0.01	0.61	0.01	0.45	0.01	PEP/E4P	[xx**3**]·[**12**xx]
Arg β (s)	0.08	0.01	0.10	0.01			0.02	0.01	0.02	0.01	AKG	[x2**3**4x]
Arg β (d)	0.71	0.01	0.69	0.01			0.74	0.01	0.79	0.01	AKG	[x**23**4x] + [x2**34**x]
Arg β (t)	0.22	0.01	0.21	0.01			0.24	0.01	0.19	0.01	AKG	[x**234**x]
Arg δ (s)	0.13	0.01	0.13	0.01	0.14	0.01	0.09	0.01		0.01	AKG	[xxx4**5**]
Arg δ (d)	0.87	0.01	0.87	0.01	0.86	0.01	0.91	0.01		0.01	AKG	[xxx**45**]
Pro α (s)	0.55	0.01	0.50	0.01	0.51	0.01	0.19	0.01	0.20	0.01	AKG	[xx3**45**]
Pro α (d1)	0.01	0.01	0.05	0.01	0.03	0.01	0.02	0.01	0.03	0.01	AKG	[xx**345**]
Pro α (d2)	0.44	0.01	0.37	0.01	0.41	0.01	0.68	0.01	0.66	0.01	AKG	[xx3**45**]
Pro α (dd)	0.00	0.01	0.08	0.01	0.05	0.01	0.11	0.01	0.11	0.01	AKG	[xx**345**]
Pro γ (s)	0.47	0.01		0.01	0.51	0.01	0.13	0.01	0.18	0.01	AKG	[xx3**45**]
Pro γ (d)	0.43	0.01		0.01	0.41	0.01	0.65	0.01	0.65	0.01	AKG	[xx**345**] + [xx3**45**]
Pro γ (t)	0.09	0.01		0.01	0.08	0.01	0.22	0.01	0.18	0.01	AKG	[xx**345**]
Asp α (s)	0.15	0.01	0.14	0.01	0.16	0.01	0.07	0.01	0.06	0.01	OAA	[12**3**x]
Asp α (d1)	0.01	0.01	0.01	0.01	0.01	0.01	0	0.01	0.01	0.01	OAA	[**123**x]
Asp α (d2)	0.12	0.01	0.11	0.01	0.12	0.01	0.24	0.01	0.24	0.01	OAA	[1**23**x]
Asp α (dd)	0.73	0.01	0.73	0.01	0.71	0.01	0.69	0.01	0.69	0.01	OAA	[**123**x]
Asp β (s)	0.57	0.01	0.59	0.01	0.53	0.01	0.2	0.01	0.25	0.01	OAA	[x2**3**4]
Asp β (d1)	0.38	0.01	0.38	0.01	0.42	0.01	0.66	0.01	0.73	0.01	OAA	[x**23**4]
Asp β (d2)	0.03	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.00	0.01	OAA	[x2**34**]

Cross peak	W3110-D		W3110-E		W3110-F		ptsG-E		ptsG-F			
(multiplet)	Intensity	SD	Intensity	SD	Intensity	SD	Intensity	SD	Intensity	SD	Precursor	Isotopomer
Asp β (dd)	0.03	0.01	0.00	0.01	0.03	0.01	0.13	0.01	0.01	0.01	OAA	[x**234**]
Thr α (s)	0.14	0.01	0.15	0.01	0.13	0.01	0.05	0.01	0.07	0.01	OAA	[1**2**3x]
Thr α (d1)	0.01	0.01	0.11	0.01	0.00	0.01	0	0.01	0.00	0.01	OAA	[**12**3x]
Thr α (d2)	0.12	0.01	0.01	0.01	0.10	0.01	0.23	0.01	0.24	0.01	OAA	[1**23**x]
Thr α (dd)	0.73	0.01	0.73	0.01	0.77	0.01	0.7	0.01	0.69	0.01	OAA	[**123**x]
Thr β (s)	0.53	0.01	0.53	0.01	0.53	0.01	0.16	0.01	0.19	0.01	OAA	[x2**3**4]
Thr β (d)	0.41	0.01	0.41	0.01	0.42	0.01	0.68	0.01	0.70	0.01	OAA	[x**23**4]+[x2**34**}
Thr β (t)	0.06	0.01	0.06	0.01	0.05	0.01	0.16	0.01	0.11	0.01	OAA	[x**234**]
Thr γ2 (s)	0.77	0.01	0.53	0.01	0.53	0.01	0.73	0.01	0.77	0.01	OAA	[xx3**4**]
Thr γ2 (d)	0.23	0.01	0.41	0.01	0.42	0.01	0.2665	0.01	0.23	0.01	OAA	[xx**34**]
Ile α (s)	0.26	0.01	0.33	0.01	0.28	0.01	0.27	0.01	0.26	0.01	OAA/Pyr	[**12**xx]·[x2x]
Ile α (d1)	0.63	0.01	0.59	0.01	0.60	0.01	0.54	0.01	0.56	0.01	OAA/Pyr	[**12**xx]·[x2x]
Ile α (d2)	0.03	0.01	0.08	0.01	0.04	0.01	0.06	0.01	0.05	0.01	OAA/Pyr	[1**2**xx]·[x**2**x]
Ile α (dd)	0.07	0.01	0.00	0.01	0.08	0.01	0.12	0.01	0.13	0.01	OAA/Pyr	[**12**xx]·[x**2**x]
Ile γ1(s)	0.79	0.01	0.79	0.01	0.85	0.01	0.65	0.01	0.70	0.01	Pyr/OAA	[x2x]·[xx**34**]
Ile γ1(d)	0.16	0.01	0.16	0.01	0.12	0.01	0.3	0.01	0.27	0.01	Pyr/OAA	[x**2**x]·[xx**34**] + [x2x]·[xx**34**]
Ile γ1(t)	0.04	0.01	0.05	0.01	0.03	0.01	0.06	0.01	0.02	0.01	Pyr/OAA	[x**2**x]·[xx**34**]
Ile δ (s)	0.55	0.01	0.55	0.01	0.59	0.01	0.79	0.01	0.79	0.01	OAA	[xx3**4**]
Ile δ (d)	0.45	0.01	0.45	0.01	0.41	0.01	0.21	0.01	0.21	0.01	OAA	[xx**34**]
Lys β (s)	0.50	0.01	0.55	0.01	0.54	0.01	0.15	0.01	0.17	0.01	OAA/Pyr	½{[x2**3**4] + [x2**3**]·[xxx4]}
Lys β (d)	0.44	0.01	0.45	0.01	0.46	0.01	0.68	0.01	0.69	0.01	OAA/Pyr	½{[x**23**4] + [x2**34**] + [x**23**]·[xxx4] + [x2**3**]·[xxx**4**]}
Lys β (t)	0.06	0.01	0.00	0.01	0.00	0.01	0.18	0.01	0.14	0.01	OAA/Pyr	½{[x**234**] + [x**23**]·[xxx**4**]}
Lys γ (s)	0.58	0.01	0.55	0.01	0.60	0.01	0.49	0.01	0.60	0.01	OAA/Pyr	[xx3**4**]·[xx3]
Lys γ (d)	0.43	0.01	0.36	0.01	0.32	0.01	0.38	0.01	0.33	0.01	OAA/Pyr	[xx**34**]·[xx3] + [xx3**4**]·[xx**3**]
Lys γ (t)	0.00	0.01	0.09	0.01	0.09	0.01	0.13	0.01	0.08	0.01	OAA/Pyr	[xx**34**]·[xx**3**]
Lys δ (s)	0.60	0.01	0.55	0.01	0.58	0.01	0.18	0.01	0.18	0.01	OAA/Pyr	½{[x2**3**4] + [x**23**]·[xxx4]}
Lys δ (d)	0.32	0.01	0.36	0.01	0.43	0.01	0.67	0.01	0.74	0.01	OAA/Pyr	½{[x**23**4] + [x2**34**] + [x**23**]·[xxx4] + [x2**3**]·[xxx**4**]}
Lys δ (t)	0.09	0.01	0.09	0.01	0.00	0.01	0.15	0.01	0.12	0.01	OAA/Pyr	½{[x**234**] + [x**23**]·[xxx**4**]}
Lys ε (s)	0.16	0.01	0.11	0.01	0.15	0.01	0.12	0.01	0.09	0.01	OAA/Pyr	½{[x**2**3] + [x**2**3x]}
Lys ε (d)	0.84	0.01	0.89	0.01	0.85	0.01	0.88	0.01	0.91	0.01	OAA/Pyr	½{[x**23**] + [x**23**x]}

Table B3: NADH,NADPH and ATP balance in w3110 and ptsG. Flux values are reported with respect to 100 moles of glucose consumed.

Reaction	Moles of NADH	Moles of NADPH	Moles of NADH+NAPDH	Moles of ATP	w3110 Flux	w3110 NADH	w3110 NADPH	w3110 Redox	w3110 ATP	ptsG Flux	ptsG NADH	ptsG NADPH	ptsG Redox	ptsG ATP
[glp]	0.00	0.00	0.00	-2.00	0.00	0.00	0.00	0.00	0.00	22.404	0.000	0.00	0.00	-44.81
[xyl]	0.00	0.00	0.00	-2.00	0.00	0.00	0.00	0.00	0.00	93.108	0.000	0.00	0.00	-186.22
[zwf]	0.00	2.00	2.00	0.00	2.45	0.00	4.90	4.90	0.00	11.681	0.000	23.36	23.36	0.00
[fba]	0.00	0.00	0.00	-1.00	97.61	0.00	0.00	0.00	-97.61	79.129	0.000	0.00	0.00	-79.13
[eno]	1.00	0.00	1.00	1.00	194.455	194.46	0.00	194.46	194.46	192.702	192.702	0.00	192.70	192.70
[pyk]	0.00	0.00	0.00	1.00	85.179	0.00	0.00	0.00	85.18	182.587	0.000	0.00	0.00	182.59
[ac]	-1.00	0.00	-1.00	0.50	151.223	-151.22	0.00	-151.22	75.61	167.113	-167.113	0.00	-167.11	83.56
[cs]	0.00	1.00	1.00	0.00	4.054	0.00	4.05	4.05	0.00	0.882	0.000	0.88	0.88	0.00
[mal]	-1.00	0.00	-1.00	0.00	-9.595	9.59	0.00	9.59	0.00	1.778	-1.778	0.00	-1.78	0.00
[tca]	1.00	0.00	1.00	1.00	2.621	2.62	0.00	2.62	2.62	0.00	0.000	0.00	0.00	0.00
[biomass]	0.004	-0.02	-0.01	-0.02	1334.47	4.73	-24.32	-19.59	-24.67	820.86	2.912	-14.96	-12.05	-15.17
														0.00
				Net NADH Balance		60.18					26.72			
				Net NADPH Balance		-15.37					9.28			
				Transdehydranse Activity		15.37					0.00			
				Net Redox Balance		44.81					36.01			
				Net ATP Balance		235.59					133.52			
				Net ATP available for Biomass		260.26					148.69	1.75		

Acknowledgements

I would like to take this opportunity to express my thanks to those who helped me with various aspects of conducting research and the writing of this thesis. I am grateful to my advisor Dr. J.V. Shanks who gave the opportunity to work on this project and gave valuable suggestions and advice constantly during this research.

I would like to thank my research group members: Dr. Ganesh, Dr. Vidya, Dr. Moon and Guy Sander with whom I had wonderful discussions and who helped me during the project.

I would like to thank my friends at Iowa state and Reddy, Saikat, Anup, Rahul in particular for giving providing constant motivations and support during my stay in Ames.